医療・創薬における

ビッグデータと人工知能

―現状と将来の方向―

東京医科歯科大学　データ科学推進室
東北大学　東北メディカル・メガバンク機構
田中　博

ゲノム・オミックス医療の現状

バイオテクノロジーの
急速な進展による
「ビッグデータ医療」時代の到来

医療・創薬への超大なインパクト
ビッグデータ時代の到来

(1) 次世代シーケンサ（Clinical Sequencing)を始めとする
「ゲノム/オミックス医療」における網羅的分子情報収集/蓄積
(2) **Biobank**/ゲノムコホート普及による分子・環境情報の蓄積
(3) モバイルヘルス(mHealth）によるWearable センサの連続計測による生理データの蓄積（unobstructed monitoring)

↓

コストレスで良質なデータが大量に収集可能

↓

治療医学の的確性の飛躍的進展:「精密医療」
医療の国民レベル・生涯ヘルスケアの進展

ゲノム・オミックス医療の２つの流れ

米国でのゲノム医療

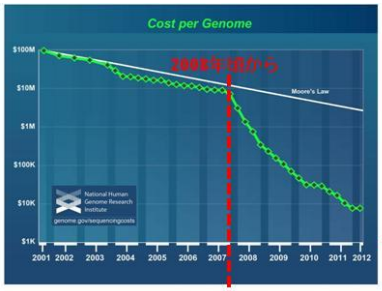
- 「シーケンス革命」(2007）からの**怒涛の展開**（2010から）
- 個々の患者の「治療医学」レベル質的向上:臨床実装の推進
  - 稀少疾患の原因遺伝子変異の同定
  - がんのドライバー遺伝子変異の同定と分子標的薬の選択
  - 薬剤代謝酵素の多型性の同定と個別化投与

欧州でのゲノム医療

- 「集合的遺伝情報」の価値⇒**ゲノム・バイオバンクへ流れ**
- 国民医療（医療の国民レベル）の向上: 社会福祉国家の理念
- 「予防医学」レベル質的向上のためにゲノム情報導入
  - 大規模前向きpopulation型バイオバンク/ゲノム・コホートの確立
  - 遺伝的素因と環境要因（生活習慣）との相互作用に基づいた「多因子疾患」の発症予測を通した「国民医療の向上」
  - 生涯的健康/疾病管理へ

TMDU

米国ゲノム・オミックス医療の流れ



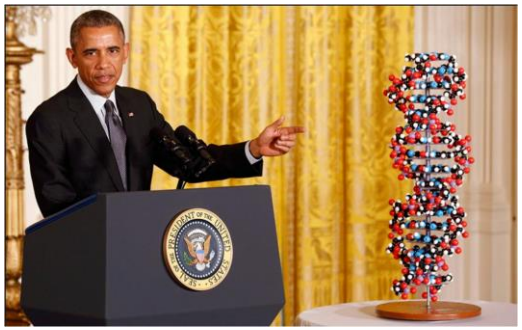
DNA Sequencing Cost: the National Human Genome Research Institute

シーケンス革命　2007/8

2005～ NGS 454 (LS,Roche)
2007/8～454, Solexa (Illumina), SOLiD (LT,TF)
シーケンス革命

	HiSeq2500		Ion Proton
本体価格	約1億円		約3500万円
モード / チップ	ハイアウトプット	ラピッドラン	Ion Proton I
解析時間	11日	27時間	2時間
リード長（bp）	2 x 100	2 x 150	200
データ産出量(Gb)	約600	約120	10
試薬コスト（ヒト1人全ゲノム）	数十万円		不可 エクソームのみ

急速な高速化と廉価化
ヒトゲノム解読計画13年,3500億円
⇒1日,10万円



オバマ大前統領　Precision Medicine Initiativeを開始、2015年1月　大統領一般年頭教書演説

先陣争いの時代

第一期

ゲノム多型性の認識
Hapmap 計画(2002)
GWAS研究など

薬剤代謝酵素の多型性の判別・電子カルテで警告・Preemptive PGx
Vanderbilt大病院

2005～ NGS 登場
(454,Solexa,SOLID)
2007/8～
シーケンス革命

Undiagnosed Genetic Diseaseの原因遺伝子POC同定
MCW小児病院

国際がんコンソーシアム開始
ICCG (2008年)
2011頃からがん変異成果報告

Cancer Driver Geneの同定と抗がん剤治験
Dana Faber CC

ゲノム・オミックス医療の臨床実装の普及
ゲノム・オミックス情報のビッグデータの出現

国家政策の時代

第二期

ゲノム医療の国家的取組み
NIH “BD2K “計画 ・各種ゲノムコンソーシアム開始

オバマ大統領　年頭教書
Precision Medicine initiative 政策の発表

精密医療普及期

第三期

100万人コホート:バンダービルト大学設開始(2016-2020)
NCI“National Cancer MoonShot” 10年計画開始
各州でのプレシジョン医療計画開始(カリフォルニア、ペンシルバニア)

2007年
2009年
2010年
2011年
2012年
2013年
2014年
2015年
2016年
2017年

TMDU

個別化医療からPrecision Medicine

個人の遺伝素因・環境要因に合わせた (tailored) 医療
One size fits for all のPopulation 医療とは異なる

趣旨：基本は、個別化医療 Personalized Medicine の概念と変わらないが診断/ 治療の個人化ではなく層別化を明確化。ゲノム一元論からの脱却

概念の革新：Personalized Medicineが標榜された時から10数年経っている

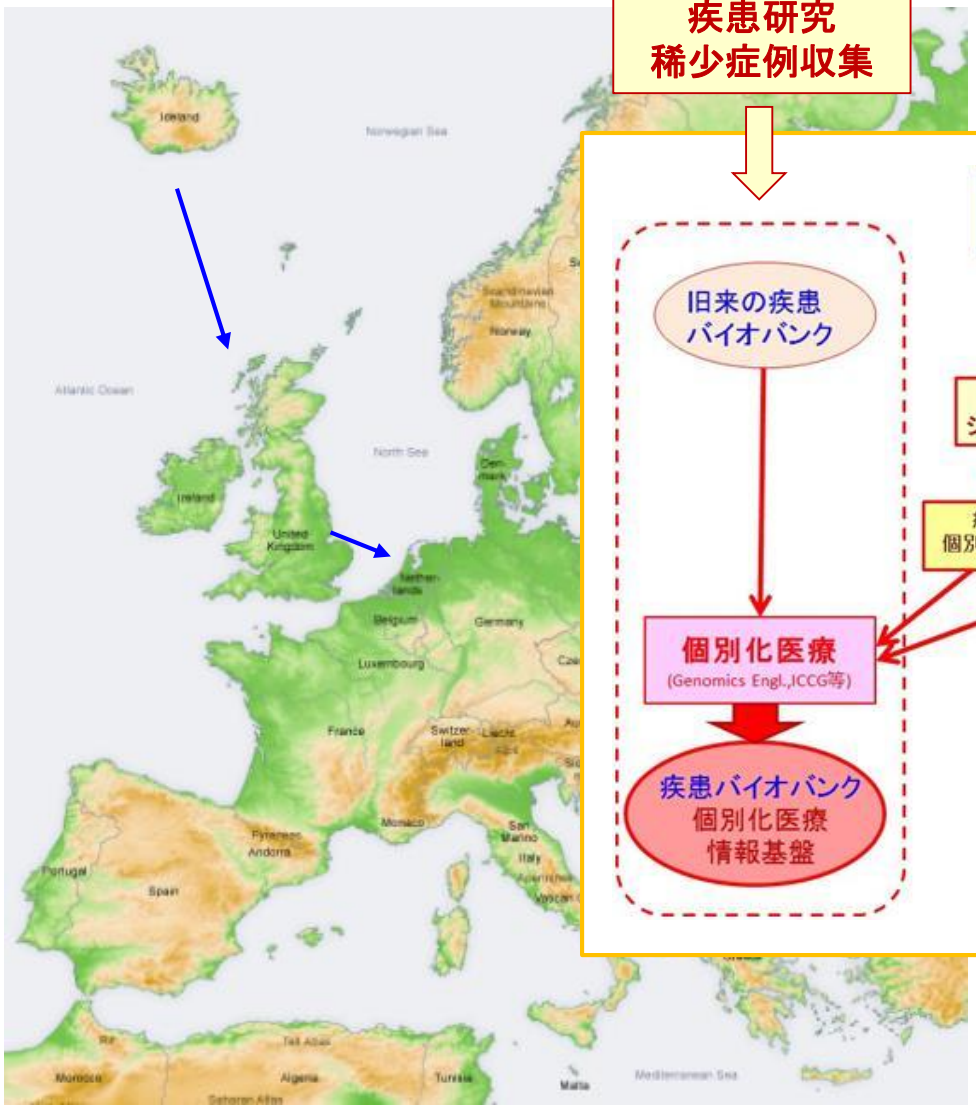
医療ビッグデータ時代の到来による個別化医療の拡張

（１）遺伝素因 X 環境(生活習慣)要因のスキーマ重視
遺伝要因(SNPや変異：Genome) だけでなく環境・生活習慣要因(Exposome) の重視
疾患発症は2つの要因の相互作用と明示的に強調。電子カルテの臨床表現型情報（Clinical Phenome)情報の重要性認識。

（２）ゲノムコホート・Biobankの重視
Precision Medicineを実現する「情報基盤」として、ゲノムコホート/Biobankが不可欠であることを認識。

（３）日常生理モニタリング情報の包摂
モバイルヘルス(mHealth)・ wearableセンサーによる大量継続生理情報の価値認識

第2の流れ　欧州のバイオバンクの普及



疾患研究
稀少症例収集

「集合的遺伝情報」による
国民レベルでの医療向上

旧来の疾患
バイオバンク

ヒトゲノム計画
(1990-2003)

次世代
シーケンス

GWAS

疾患
個別化機序

ゲノム
生得的
多型性
SNPアレイ

個別化医療
(Genomics Engl.,ICCG等)

疾患バイオバンク
個別化医療
情報基盤

Island/deCODE社
国民遺伝情報DB

前向きコホート概念

ELSI慎重対応

Population準拠

家系情報

UKバイオバンク
ゲノムコホートの確立

LifeLines
3世代コホート

東北メディカル
メガバンク

出生コホート

大規模前向きゲノムコホート
国民的医療・予防医療
情報基盤

ビッグデータ医学/医療の第２の流れ
Biobankとゲノムコホートの世界的興隆

バイオバンクの目的･機能の変化

- **従来は**：稀 少疾患組織標本や臨床研究の**資料保存**のため
- **最近は**：**ビッグデータ時代での変換**：**ゲノム医療の情報基盤**としての役割・**世界的に普及**
- **疾患BioBank**：**ゲノム・オミックス個別化医療／創薬**の情報基盤：
  - 従来の疾患バイオバンクが**個別化医療の概念により変革、個別化医療の情報基盤**としての役割
  - **疾患罹患患者の網羅的分子情報**とそれに対応する**臨床表現型情報**の収集。
  - **疾病の個別化分子機序解明や治療戦略構築、予測医学・創薬科学への貢献**
- Population**型BioBank**：**「集合的遺伝情報」国民医療レベルの向上、予防医学**の情報基盤
  - 「健常者」前向きコホート。網羅的分子情報（genome）と環境情報（exposome）収集
    長期間（生涯）を追跡するゲノム・コホート
  - 主に遺伝子素因情報も含めた**「多因子疾患」**の**疾患の発症リスク予測、重症化予測**

欧米のBiobank

- **英国　UK biobank**
  - 50万人の健常者。40～69歳（2006-2010, 62Mポンド), 追加調査（2011-16, 25Mポンド）
  - 健診データ（血液・尿・唾液サンプル、生活情報）とゲノム情報（SNPアレイを集め、健康医療状況を追跡する。アイスランド、deCODE社の「国民遺伝子情報データベース」の概念を新たに実現
- **英国　Genomics England**,
  - 2013開始、2017年までに 10万人のゲノム 配列収集。全ゲノム次世代シーケンス
  - 最初の対象は稀少疾患（患者・家族）、がん患者、最初はEnglandのみ。企業との協同重視
- **欧州　BBMRI**（Biobanking and Biomole Research. Infrastructure.)
  - 250以上の欧州各国のBioBankを連携する
- **オランダ　Lifeline**
  - 165000人北部オランダ　2006年開始　30年間の追跡、家系情報・3世代コホートを世界初で実現
- **米国　Precision Medicine Initiative, Genome Cohort：“All of Us”コホート**
  - これまでのBiobank（例えばBioVUなど）を集めて100 万人のゲノムを集める

TMDU

ビッグデータ医学/医療の２つの流れに起因する大規模な生命情報DB/KBの出現と利用

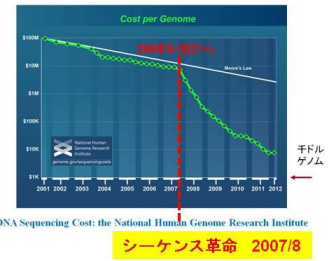
- ヒトゲノム解読計画以降急速に進展
  - Hapmapプロジェクト,1000 genome,がんICGC,TCGA、TopMED
  - ゲノム変異・多様体
    - dbSNP, HGMD, **Clinvar**, **Clingen**, OMIM,GWAS catalog
    - 表現型との対応: dbGaP, EGA
  - 遺伝子発現プロファイル
    - 疾患特異的transcriptome: **GEO, ArrayExpress**,
    - 薬剤特異的trancriptome: **c-Map, LINCS**
  - タンパク質
    - ３次元構造：PDB, Swiss-Prot,
    - タンパク質間相互作用：**HPRD, STRING,** BIND
  - 分子ネットワーク、パスウェイ
    - KEGG, TRANSFAC, BioCyc, Reactome
- 各種バイオバンク症例ベース（制限アクセス）
  - UK biobank, BMBRI, 東北メディカル・メガバンク
- これらの大規模DB/KBを組合せてゲノム医療/創薬を推進

TMDU

医学/医療へのビッグデータの衝撃

次世代シーケンサの登場

シーケンス革命（2007)



コストレスで高精度な網羅的分子情報の出現

1. ゲノム・オミックス医学/医療の進展
   - Clinical Sequencingによるゲノム・オミックス医療の臨床実装の急速な進展
2. Biobank/ゲノムコホートの世界的普及
   - 個別化医療/予防の情報基盤として普及
3. 大規模な生命情報DB/KBの出現
   - ゲノム・オミックスによるDB/KBの膨大化

わが国での現状「ゲノム医療元年 (2016)」

■「ゲノム医学実現推進協議会」(中間報告）2015.7

研究費を用いた試行的ゲノム医療であるが、いくつかの医療施設で
ゲノム・オミックス医療が試行されていた

- 例：がんの網羅的分子診断と個別化治療
  - 国立がん研究センター（Top-gear、SCRUM-Japan)
    - ドライバー遺伝子の診断。分子標的薬の治験グループに割当て
    - がんのゲノムパネル：来年先進医療（7施設）
  - 岡大,京大, 北大,千葉大 病院併設型BB
- 予定：2018年度より「がんゲノムパネル」先進医療（7か所）開始

■AMED（日本医療研究開発法人）がゲノム医療を推進に予算

- IRUD（Initiative on Rare and Undiagnosed Disease)
  未診断疾患の原因遺伝子をIRUD拠点病院が審査して解析センターがシーケンシング。その後、DB化する。
- ゲノム医療実現推進プラットフォーム事業
- 臨床ゲノム情報統合DB事業

ゲノム医療臨床実装では、米国と水を空けられている。しかし、
Biobank/Genomic Cohortでは我が国の状況は遅れていはいない。
米国とは異なったBiobank準拠のゲノム医療/創薬を推進するべき

TMDU

ビッグデータ医療の課題

医療の「新しいビッグデータの革命性」

～ゲノム・オミックスデータの基軸的な特徴～

＜**目的**も**データ特性**も従来型と違う＞

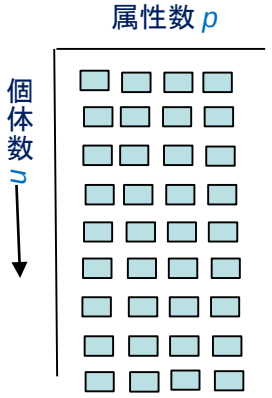
従来の医療情報の「ビッグデータ」(n≫p)

医療情報・疫学調査では **属性数**：**数十項目程度**

個体数：近年電子化の流れ⇒**個体数：膨大**

– 目的：Population （集合的）医学のBig Data

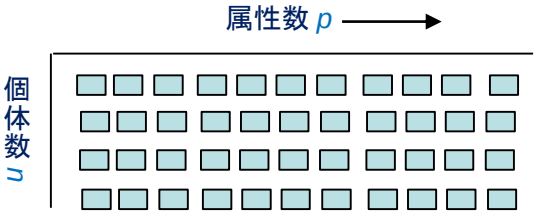
⇒個別を集めて「集合的法則」を見る



網羅的分子情報などのビッグデータ(p≫n)

１個体のデータ属性数が膨大（SNP4000千万）

ただし個体数は大規模biobankでも数十万



属性(*p*)>> 個体数(*n*) :従来の多変量統計学が無効

「新 ***np*** 問題」: GWASは単変量解析の羅列

– 目的：医療の場合 個別化医療 Personalized Medicine

⇒大量データを集めて「個別化パターン」の多様性を摘出

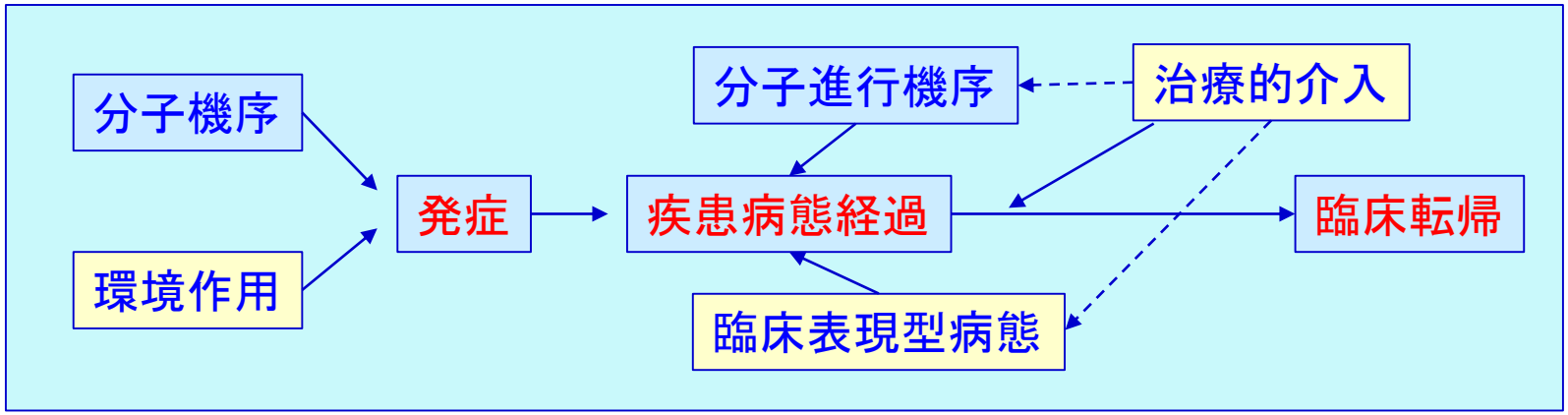
新しいデータ科学の必要性

TMDU

医療の「ビッグデータ」革命は どんな既存のパラダイムに変革しているか

- Population （集合的）医学からパラダイム転換
  - <One size fits for all>の集合的医療はもはや成り立たない
  - 個別化医療 “Personalized medicine”の概念
    - 個別化医療実現のために＜個別化・層別化パターン＞がどれだけ有るか網羅的に調べる：どこまでの粒度で個別化・層別化すればよいか
- Clinical research（臨床研究）のパラダイム転換
  - 臨床研究の基礎：従来の範型RCTは,個別化概念を取扱えない
  - <EBM: (statistical) evidence based>の呪縛からの解放
  - 「標本」統計・「推測」統計学に制約されない臨床研究
  - Real World Data・ビッグリアルワールドデータからの知識生成
  - Learning Health System: 学習的医療実践

課題１対応する非ゲノム病態データの検証的「情報化」



疾患経過のオントロジー

疾患分子発症進行機序（生体分子ネットワーク）

対応する非分子機序の明確化

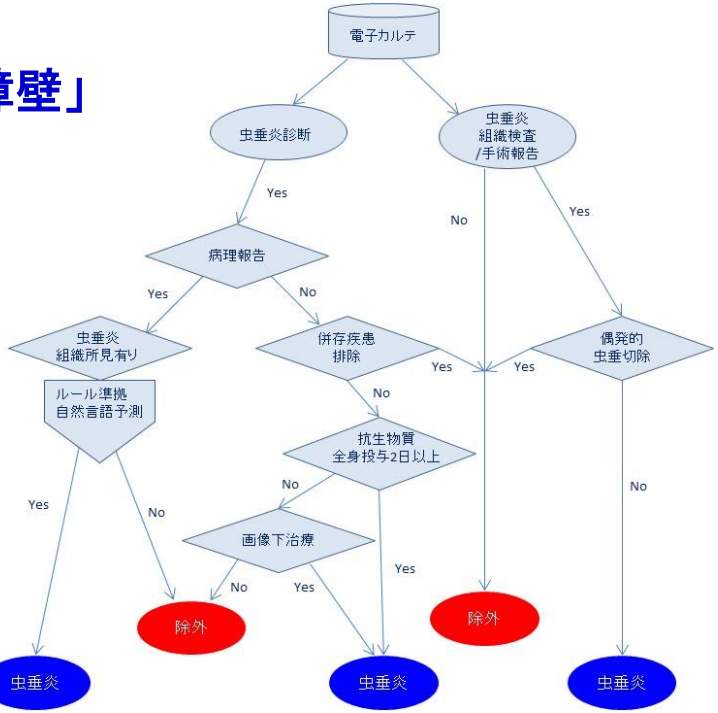
環境発症要因　臨床表現型情報　治療介入効果

臨床表現型との統合(phenotyping)

臨床表現型データ検証的抽出、「非構造化データ障壁」

electronic MEdical Records + GEnomics (NHRI-funded

- **phase I** (2007-2011) EMR-bsedゲノム研究の探求
  - EMR(臨床phenotyping)とbiorepositoryに基づくGWAS等（EMR-based GWAS）が可能か（LHS）。
    - 開始時はGWAS全盛時代。ゲノム医療の臨床実装未着手
  - 電子カルテより臨床表現型情報抽出 phenotypingルール
  - 計画開始時参加施設：Mayo, Vanderbilt Univ, Marshfield, Univ. Washington, Northwestern Univ.など５施設,
- **phase II** (2011-2015) 臨床実装へ舵を切る
  - MCWの臨床実装のインパクト,Vanderbiltの先制PG x
  - 電子カルテと遺伝情報の統合
    - 電子カルテへのゲノム情報の統合
    - **PheKB** (Phenotype Knowledge Base)
    - ゲノム医療の実装、PGxの臨床応用
    - 結果回付 **Return of Result,** ELSI等
  - 4つのサイトが新しく加わる
    - 小児病院グループとMount Sinai, Geisinger
- phase III （2015より始まる
  - NHGRIのコンソーシアムと連携
  - とくに**CSER** “Clinical Sequencing Exploratory Research”



PheKB: phenotyping ルール

課題２生命医療情報のビッグデータ化による「革新的(innovative)知識」発見の困難性

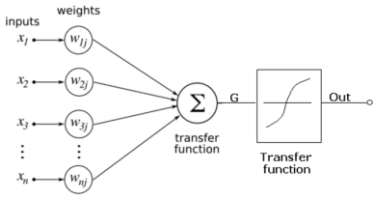
- 臨床ゲノム医学
  - 全ゲノム配列の普及、多層オミックス情報の収集、分子画像の発展
  - ビッグデータ化：超多次元相関ネットワーク
  - 〈網羅的分子情報と臨床表現型情報〉の相関
- 予防ゲノム医学
  - バイオバンクの大規模化、国際連携によるバーチャル連携
  - 〈遺伝的素因と環境/生活様式要因〉の相互作用と発症の相関ネットワーク

いずれも超多次元複雑ネットワークの縮約理論

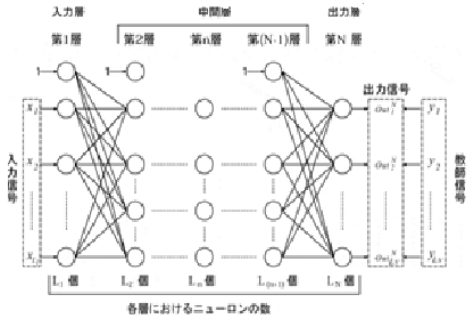
TMDU

人工知能 Deep Learning への期待

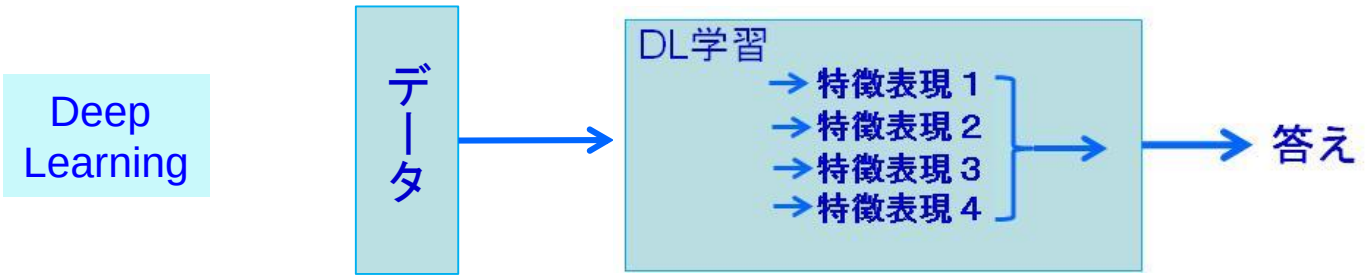
- 機械学習のこれまでの限界
  - 「教師あり学習」
    - 分類対象の特徴と正解を与え学習機械（AI）を構築
- Deep Learningの革命性
  - 「教師なし学習」
    - 対象の特徴表現や対象の高次特徴量を自ら学ぶ



神経情報素子

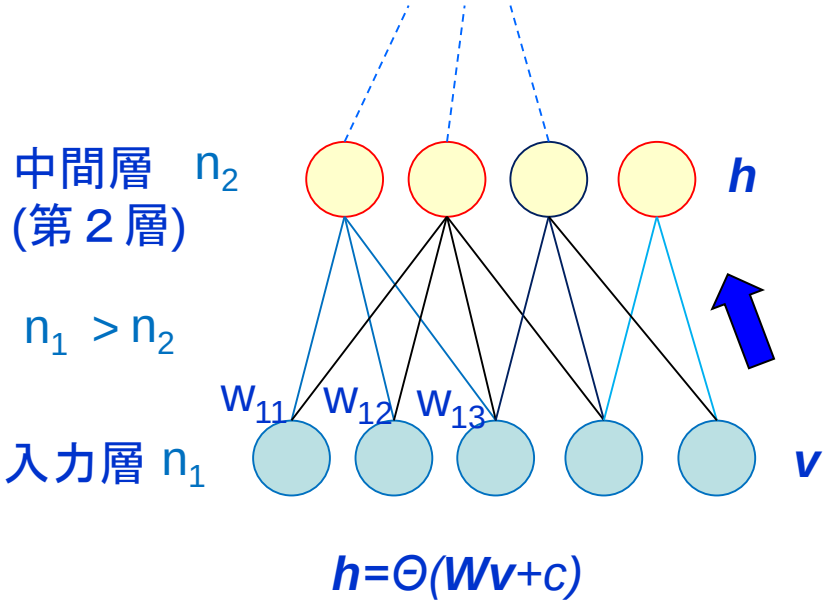


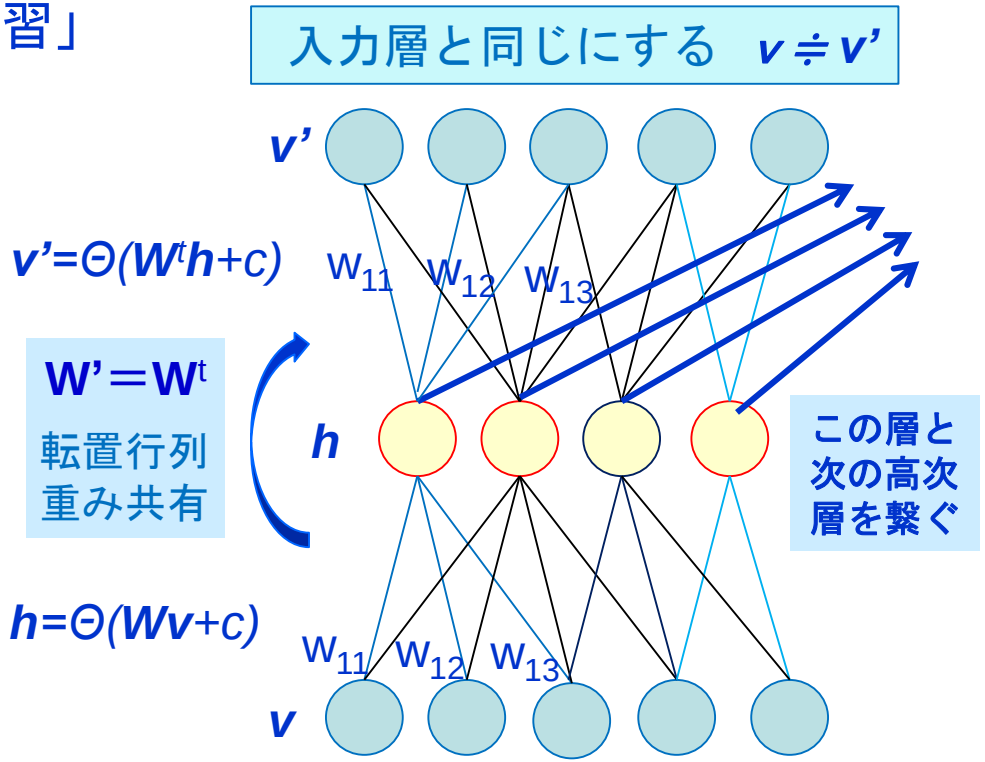
多層ニューロネットワーク



DLの革命点 Autoencoder

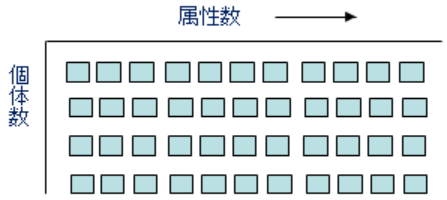
- 対象に固有な内在的特徴を学ぶ自己符号化の原理
- 格段ごとに入力の少ない中間層を入力へ逆投影して復元できるか
- 次元を圧縮され可及的に復元する（1000 nodes ⇒100nodes＝?⇒1000nodes）
  →できるだけ復元に効果的な特徴量を探索する
  →内在的な特徴量を見出す
- 最終層で人との対応「教師あり学習」





「ビッグデータ」のData 縮約原理

問題点　　属性項目数(p)≫サンプル数(n)



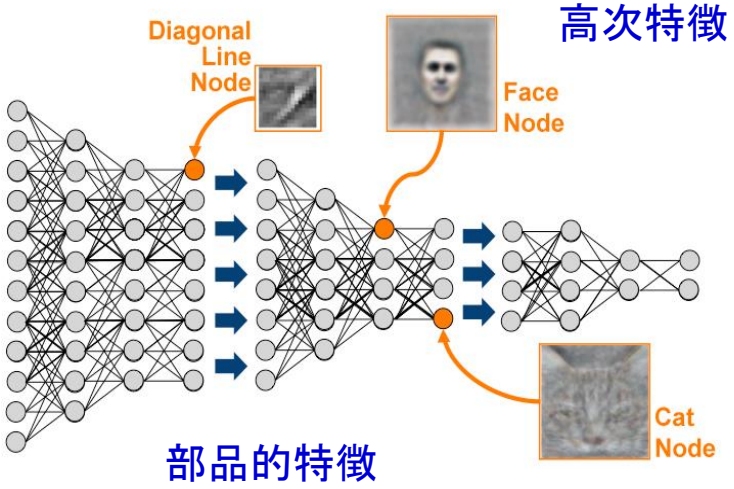
p: 数億になる場合あり、n: 多くても数万

これら膨大な属性変数がすべて独立ならばビッグデータの構造解析は不可能。単変量解析の羅列（GWASのManhattan Plot)しか可能でない

ビッグデータ・スパース仮説

ビッグデータは、多数であるが属性値数より少ない独立成分が基底となって、相互にModificationして構成されている。
「構造」の存在：推定した結果で判明

データ構成性の原理
principle of compositionality



構成性の推定（Google猫の例）

Deep Learningによる
多次元ネットワーク縮約法
(Hase, Tanaka 2017)

- 医療・創薬ビッグデータへの応用性は高い
- 超多次元ネットワーク情報構造の急増
  - ゲノム医療＜網羅的分子情報ー臨床表現型情報＞
  - ゲノムコホート＜遺伝素因ー環境要因(生活習慣)＞
- Deep Learning-based Network Contraction
  「DLネットワーク縮約法」

  超多次元ネットワーク情報構造⇒
  少数の特徴的ネットワーク基底に分解
- 線形分解ではない。非線形分解で基底への射影
  - 線形分解（特異値分解：SVD）との比較

TMDU

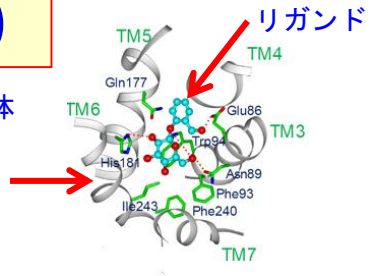
人工知能応用としての
AI創薬

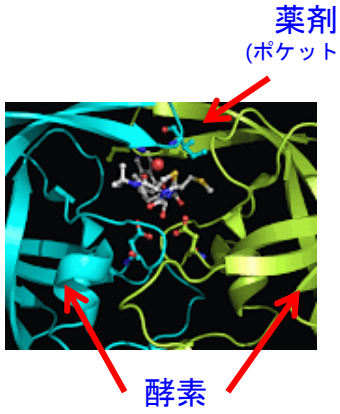
TMDU

ビッグデータ計算創薬１

計算創薬(computational drug discovery) の新しい方向

これまでの計算創薬 (*in silico* 創薬)

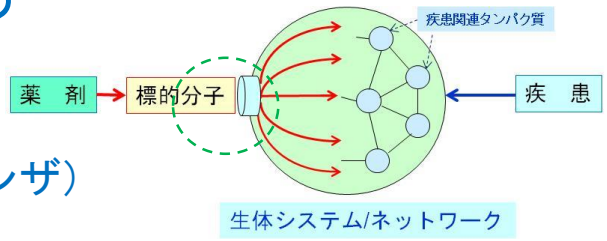




分子(結合構造)中心

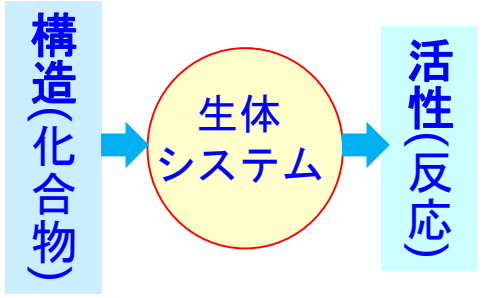
- **分子構造解析・分子設計**
- **Structure-based rational drug design**
- **標的分子**(受容体・酵素)**と薬剤**(リガンド)との**結合構造**（ポケット）の分子構造を根拠に
- **リガンドの分子設計(量子化学等を使う)**
  - 成功例：インフルエンザ薬　ザナミビル（リレンザ）
- 標的に結合する**リード化合物・構造最適化**
- **結合後の生体システム**の反応・振舞い

  ⟹ 明確な取扱いがない



定量的構造活性相関(QSAR)

- 化合物の分子**構造**と生体**活性**の**関係**
- **しかし両者の間**には**生体システム**がある

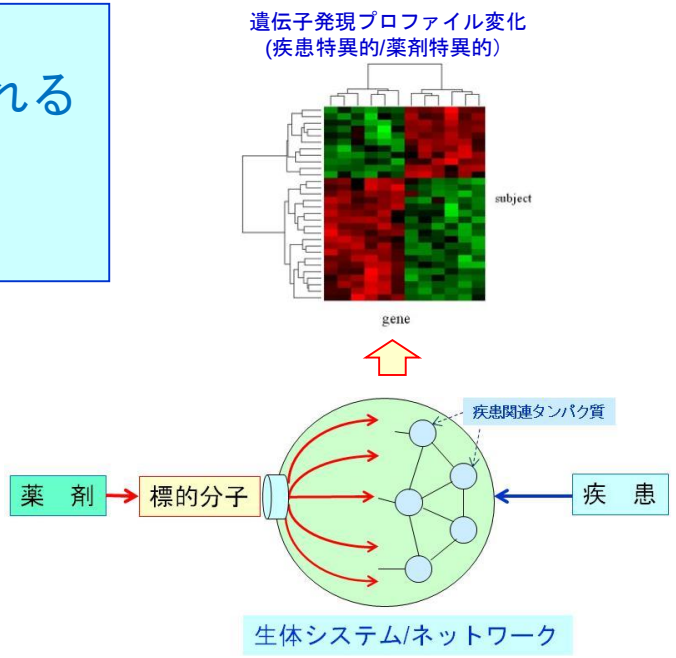


ビッグデータ計算創薬２

新しい計算論的創薬のアプローチ（生体分子プロファイル型創薬）

疾患罹患状態における
疾患関連遺伝子（タンパク質）に起因し決定される
疾患時の**生体のゲノムワイドな特異状態**
疾患特異的な網羅的分子プロファイル変化

薬剤投与による
標的分子と薬剤分子の結合に起因し起こる
投与時の生体のゲノムワイドな反応/振舞い
薬剤特異的な網羅的分子プロファイル変化



網羅的分子プロファイル⇒分子ネットワーク全体変化

＜疾患状態の生体＞に＜薬剤ー標的分子の結合＞から引き起される作用によって
ゲノムワイドな 生体分子環境がどう変化するか「生命システム観点からの理解」

化合物, 標的分子, 疾患間の関係の「ビッグデータ」DBを利用

TMDU

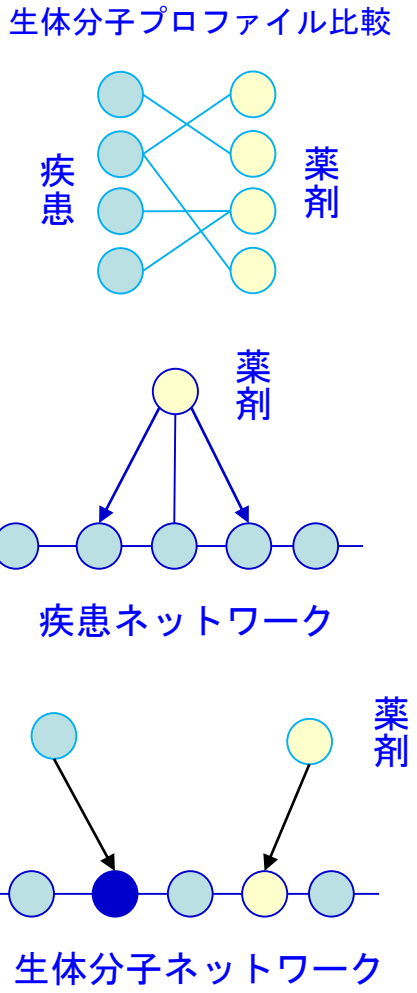
生体分子プロファイル型創薬/DR 方法論の深化

深化

第１段階：疾患・薬剤プロファイル直接比較
- 疾患罹患時と薬剤投与時の生体反応の遺伝子発現プロファイルを比較。
- パターン正負相関性に基づく有効性毒性予測

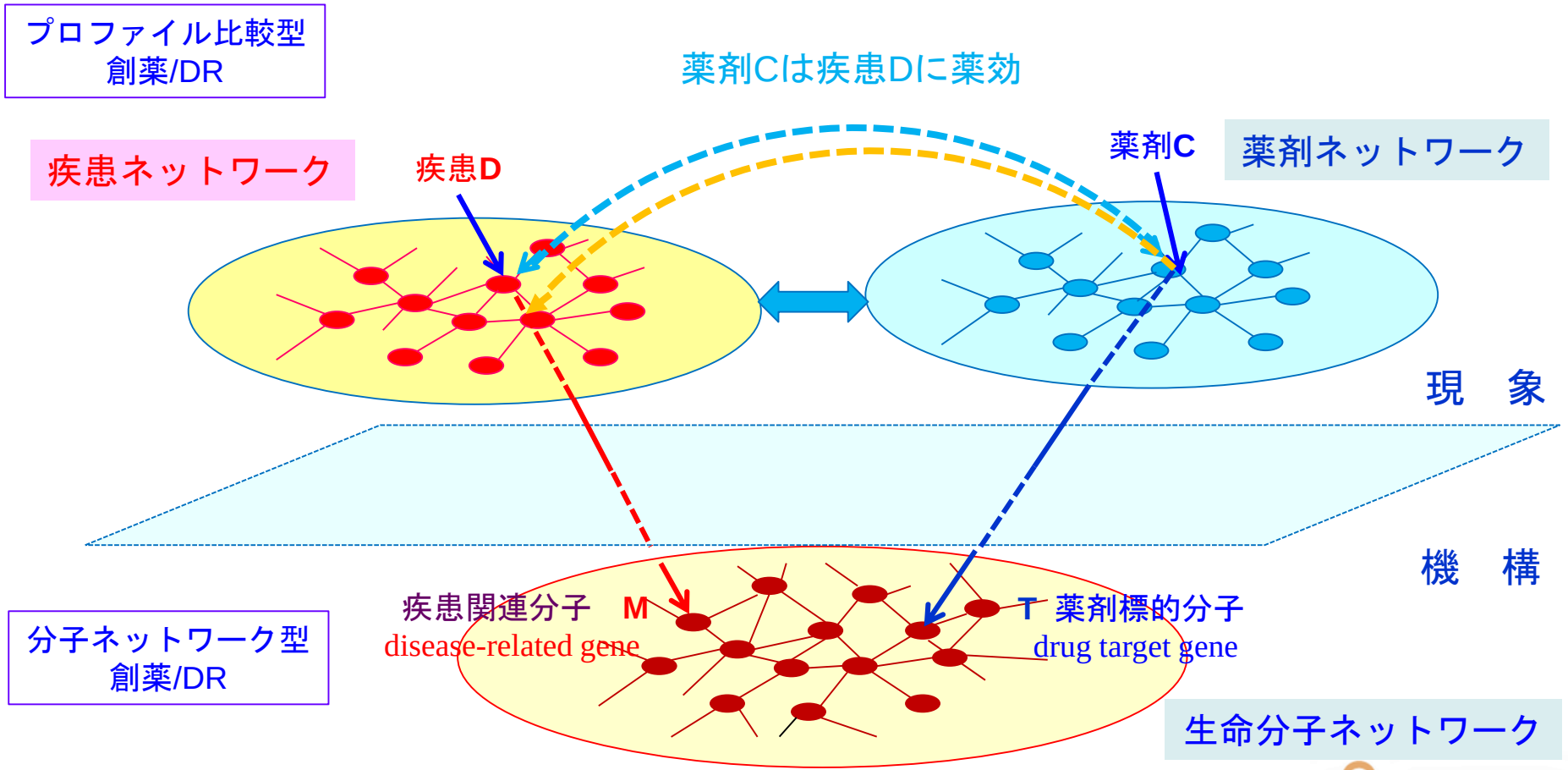
第２段階：疾患・薬剤ネットワーク近接解析
- 疾患あるいは薬剤の集合をネットワーク表現
- ネットワーク近接性に基き有効性・毒性予測

第３段階：生体ネットワーク媒介型比較
- 生体分子ネットワークを＜場＞として、疾患・薬剤の作用の足場分子を同定
- 足場分子間の相互作用（総合的距離）の評価に基づき有効性・毒性予測

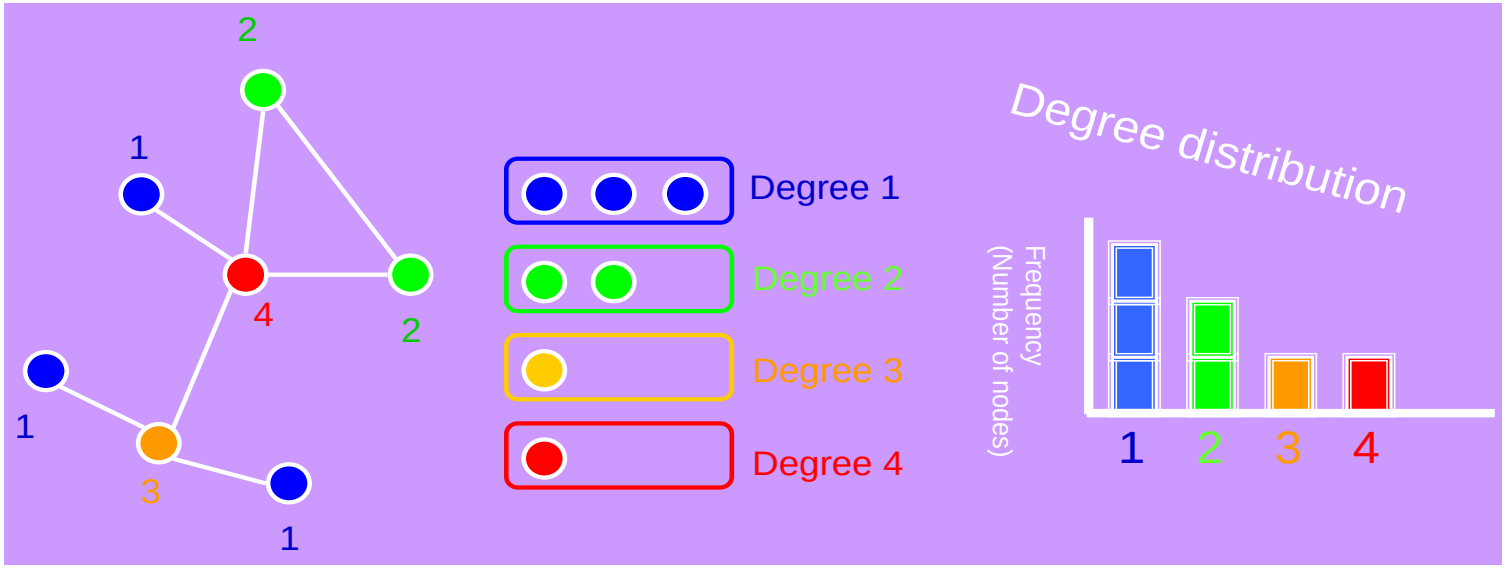


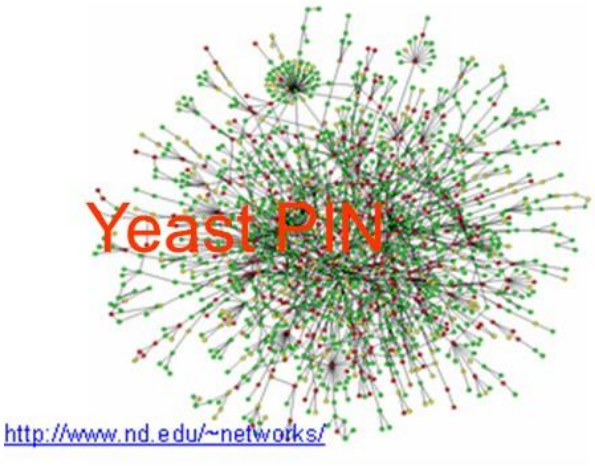
生体分子プロファイル型計算創薬/DRの基本的枠組み

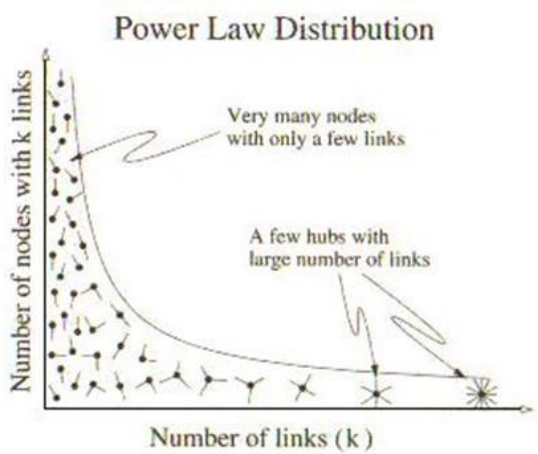
3層の生体・薬剤のネットワーク間の関係図式



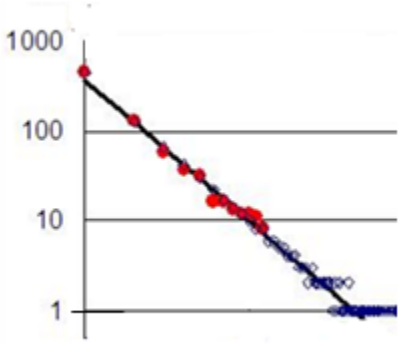
タンパク質相互作用ネットワーク(PIN)では数少ない相互作用が集中したタンパク質(hub)と相互作用が1や2の多数の末端タンパク質(branch)が存在する





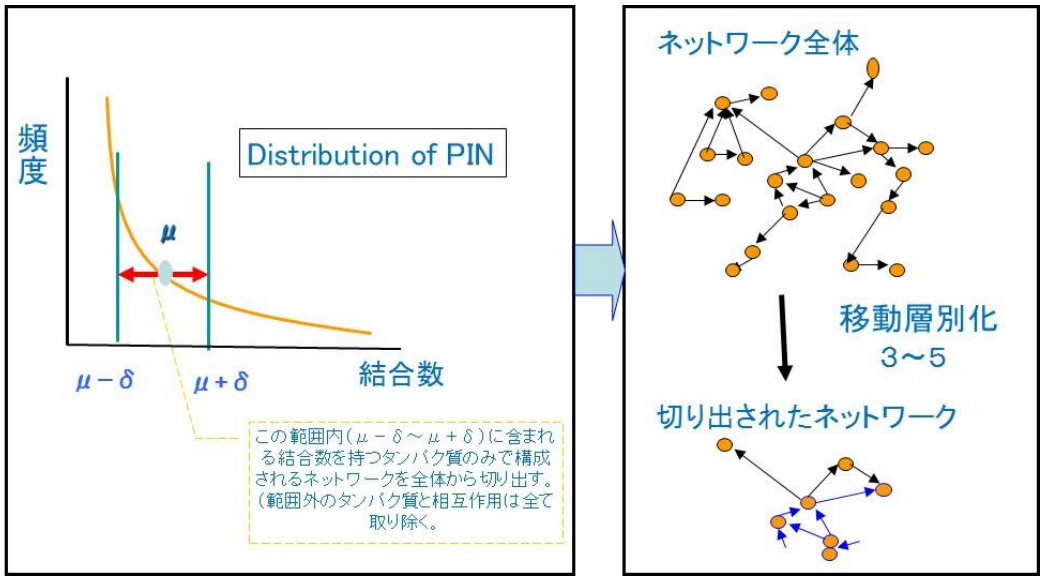


http://www.macs.hw.ac.uk/~pdw/topology/ScaleFree.html

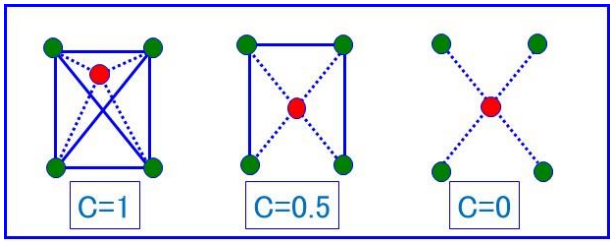


Log-log変換で直線

結合次数ごとの部分ネットワーク構造の結合密度の解析

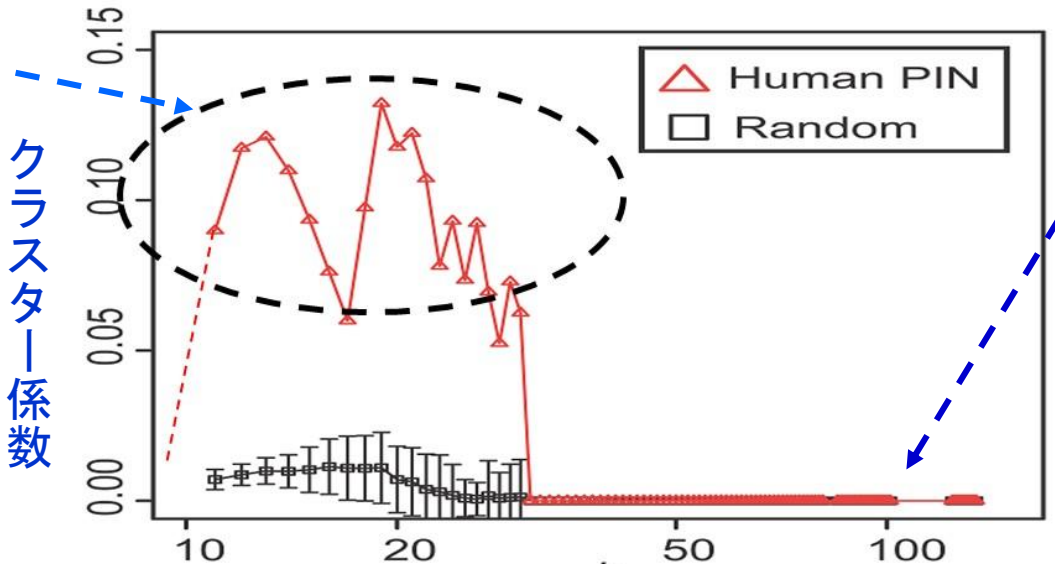


クラスター係数



Hase, T., Tanaka, H et.al (2009)
Structures of protein protein interaction network and their implications on drug design. ***PLoS Compt Biol***. 5(10):

中程度の次数（7～42）を持つタンパク質は多数の密なモジュールを構成

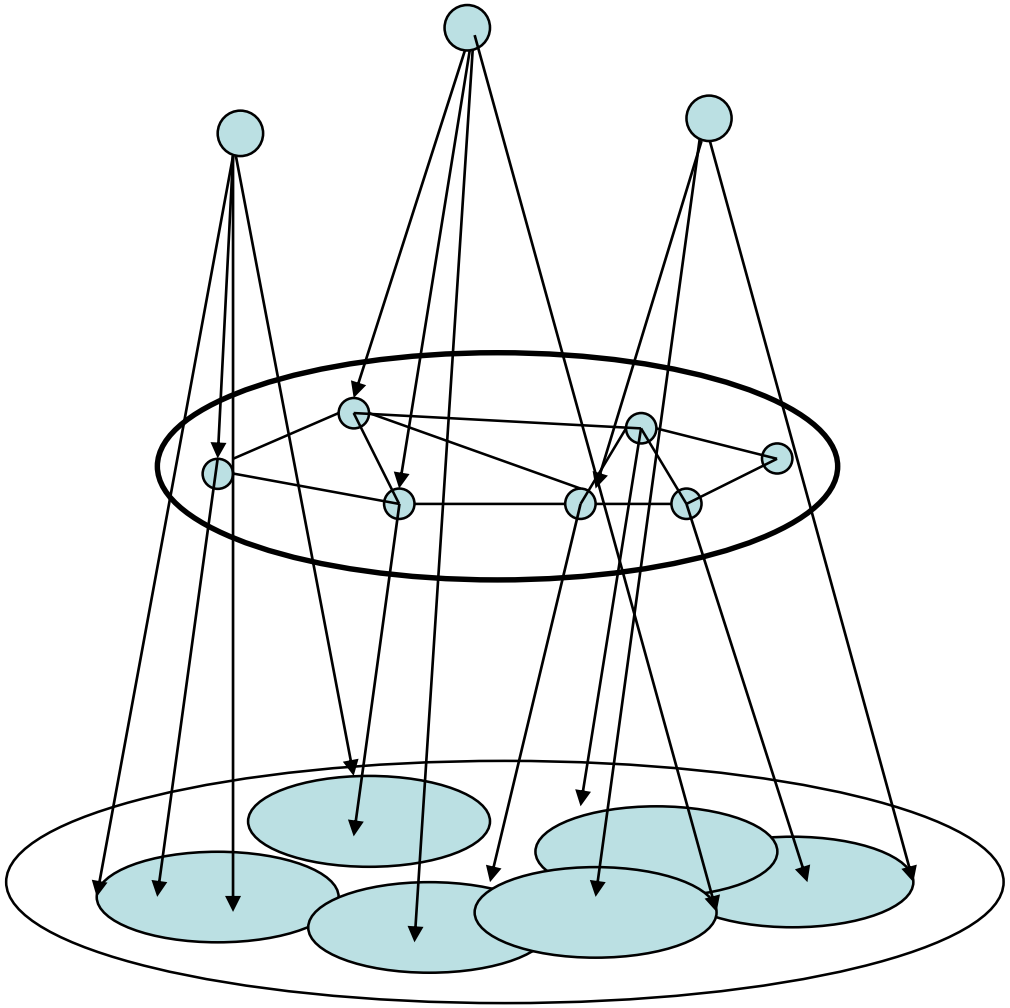


高い次数を持つノード（スーパーハブ）はお互いに密に結合しない

TMDU

タンパク質相互作用から見られる

生命情報ネットワークの構造



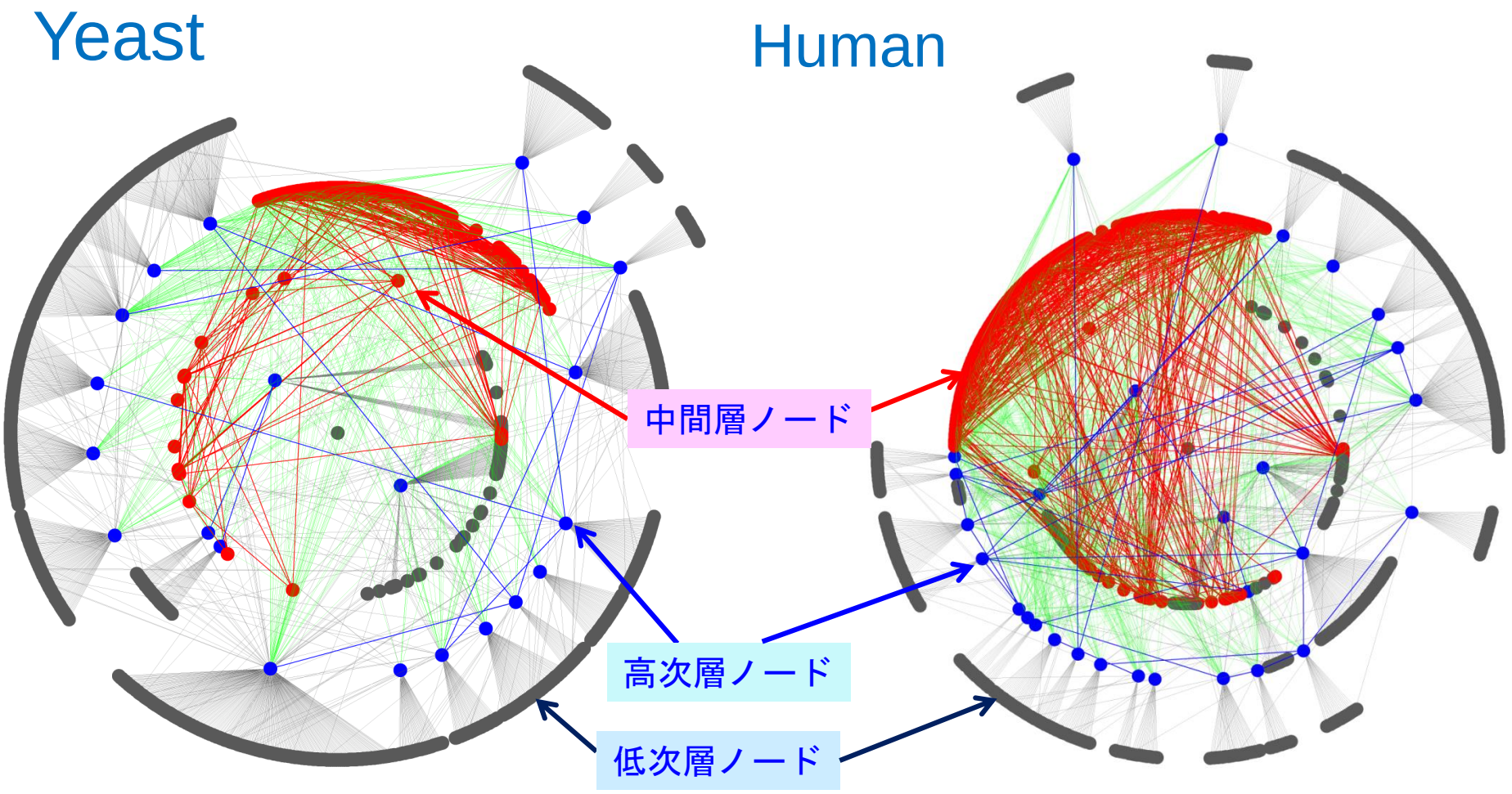
高層
高次数　ハブ
次数
＞31 ヒト
＞39 酵母

中間層
中程度次数
次数
6 ～30　ヒト
6 ～38　酵母

低層
低次数　ブランチ
次数＜6

TMDU

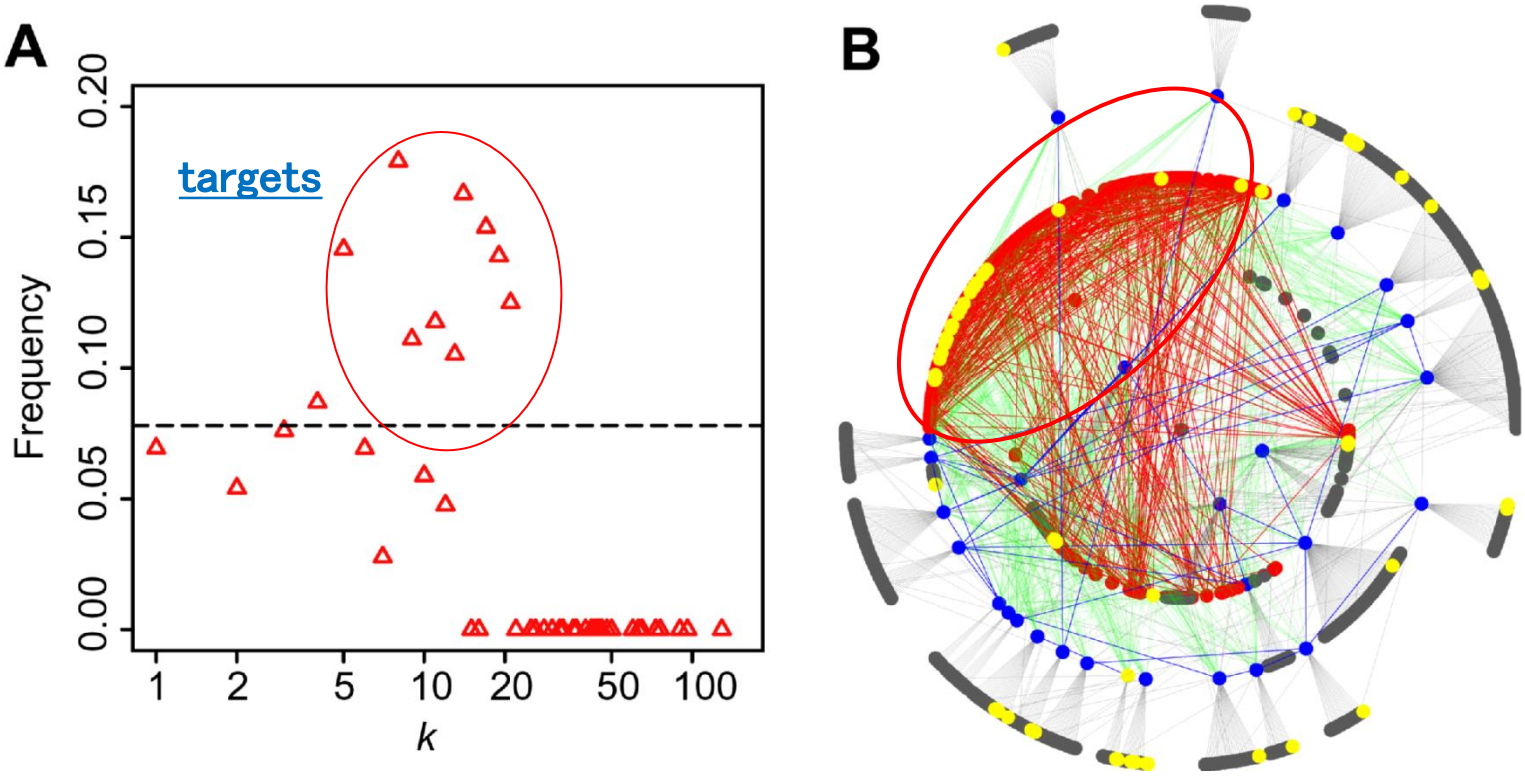
タンパク質相互作用ネットワークの Cloud Topology（3環トポロジー）



中間層の次数 ノードは PPI バックボーン を形成する

灰色, 赤, 青は、それぞれ低層、中層、高層の次数のノードをそれぞれ表す。

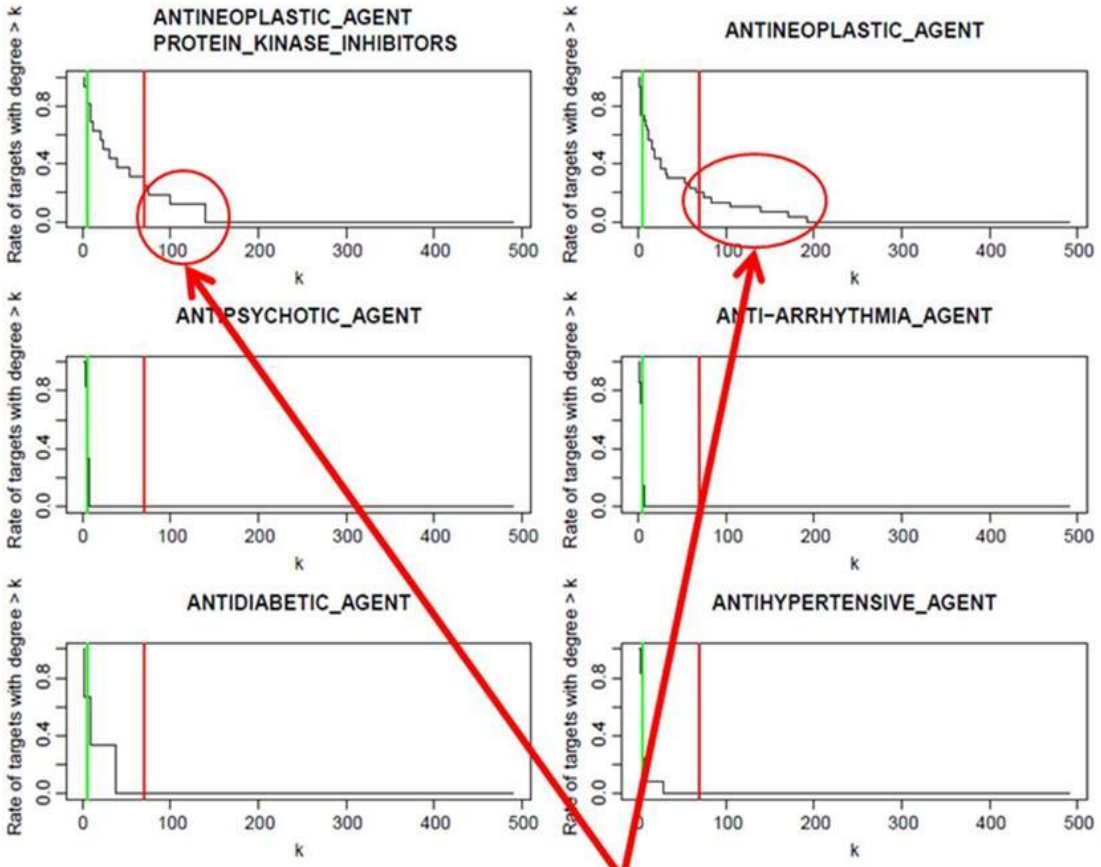
薬剤標的分子と結合度数



中層レベルのノードは治療薬として最適な標的である。それゆえ、多くの市場にある薬剤標的は、ヒトのバックボーンタンパク質に集中している

TMDU

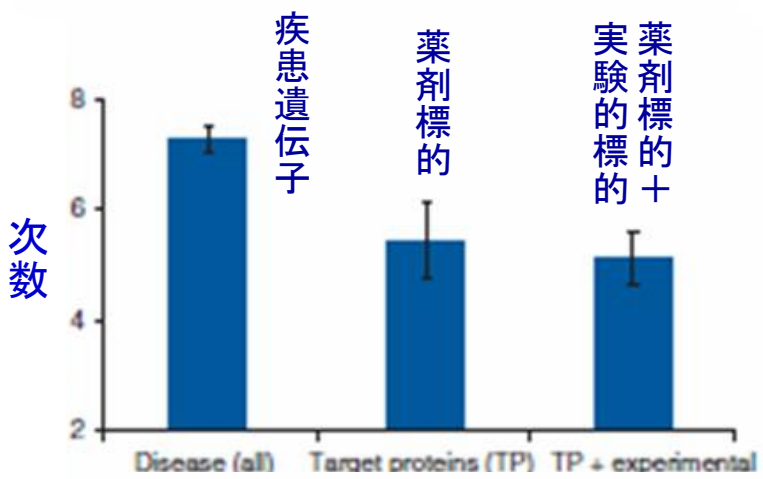
がん疾患遺伝子は高層次数ハブのタンパク質が多い



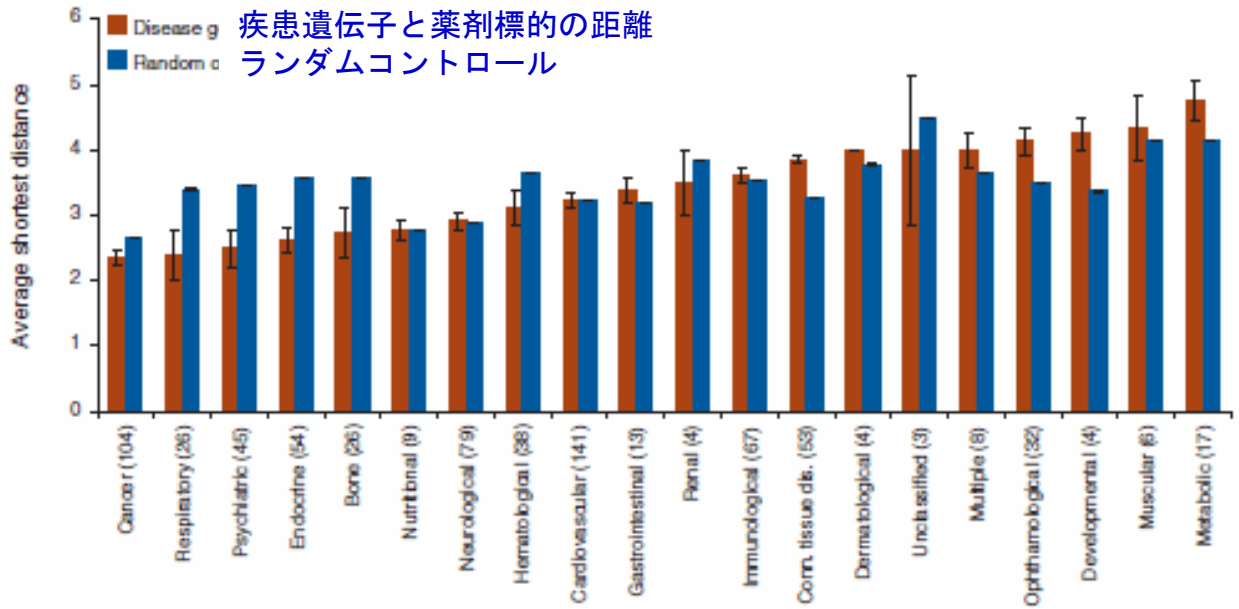
~30% のがんは抗がん剤の標的は高次のタンパク質相互作用数

標的タンパク質と疾患遺伝子の距離

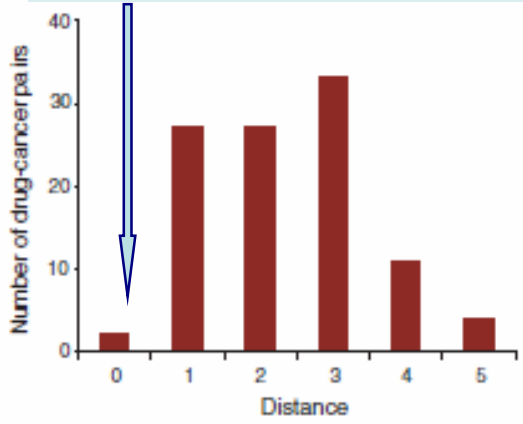
薬剤標的タンパク質と
疾患関連タンパク質の
間の距離：２～４リンク



Yıldırım　M A, et al,
NATURE Biotechnology
2009



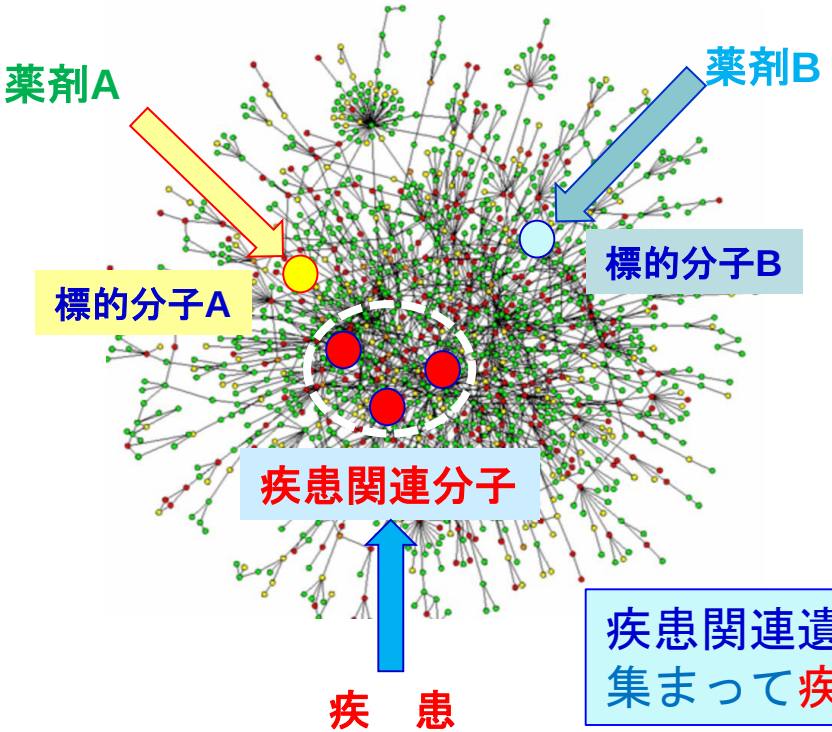
抗がん剤の場合
疾患遺伝子と距離０の標的



抗がん剤の標的分子と
疾患遺伝子の間に距離

標的分子や疾患関連分子の
タンパク質相互作用ネットワーク（PPIN）

- 薬剤ネットワークと疾患ネットワークの基盤：生体分子ネットワーク
- タンパク質相互作用ネットワーク（PPIN）での創薬/DR戦略
- PPIネットワーク場を基礎にして距離（近接性）を検討
- **薬　剤**：**薬剤の標的分子**（タンパク質）によって　PPI場と繋がる
- **疾　患**：**疾患特異的発現遺伝子**を疾患関連分子（**タンパク質**）へ翻訳、
- PPIN場内での**薬剤（標的分子）と疾患（疾患関連遺伝子）の「代理人」**の**距離・近接性**を基準に、薬理作用のインパクト力を評価

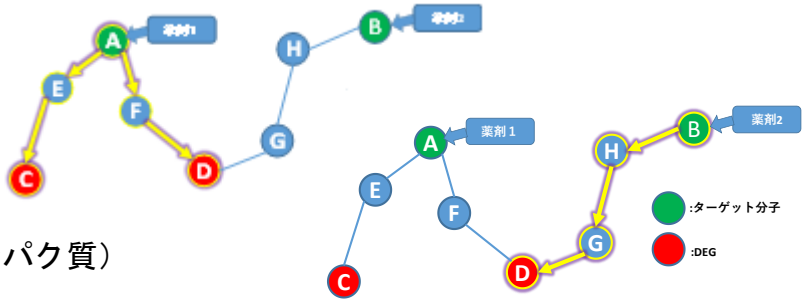


タンパク質相互作用
ネットワーク（PPIN）

疾患関連遺伝子はネットワーク上の近傍に集まって疾患モジュールを形成する

PPIの基づくDR（肺腺癌の例）

- **Interactome**(タンパク質相互作用)ネットワーク（Sun, 2016）
  - HPRD
    - 37,070 PPI, 9465 タンパク質
  - STRINGS
- **薬剤⇒標的分子**：DrugBank
  - 7,759 薬剤、4300タンパク質
  - 12,604 薬剤ー標的分子 （4,452薬剤, 1,617タンパク質）
- **疾患遺伝子（肺腺癌）**
  - TCGA (The Cancer Genome Atlas)より差異的発現遺伝子を同定
    - 445 肺腺癌例, 19 正常例, 疾患遺伝子 FC >2.0 or <0.5, 927 差異的発現遺伝子
- **薬剤の疾患遺伝子への影響力 評価IPS** (Impact power score)
  - **薬剤の標的分子**と**疾患遺伝子**の間のネットワーク距離の総合評価
    「再出発ありランダム歩行」RWRでネットワーク距離を評価
  - **標的分子からランダム歩行を繰り返す**（出発点から再出発あり）
    s時点後, 疾患遺伝子のノードにどれだけの確率で滞在しているかをIPSとする
  - 一定の時間が過ぎると、**定常状態**になり、歩行で滞在確率分布は変化しない。
    定常状態での**疾患遺伝子ノードに滞在している確率**の総和が薬剤の評価になる



$$\mathbf{P}^{s+1} = (1-\gamma)\mathbf{M}\mathbf{P}^{s} + \gamma\mathbf{P}^{0}$$

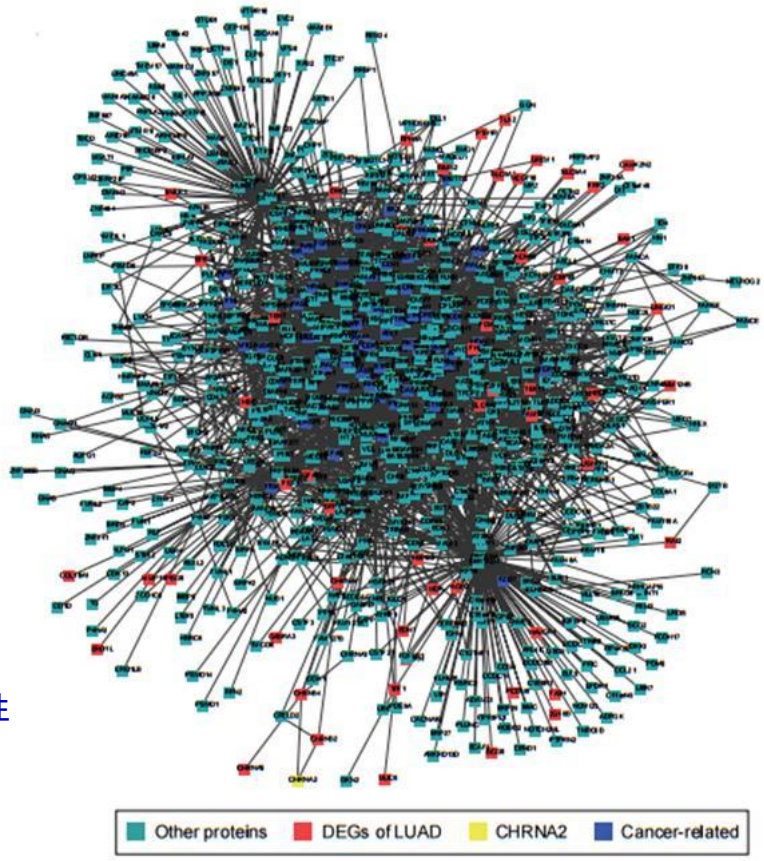
$\mathbf{P}^{s}$：時点sでの各ノードでの滞在確率　$\mathbf{M}$：各ノードへの遷移確率　γ：再出発確率

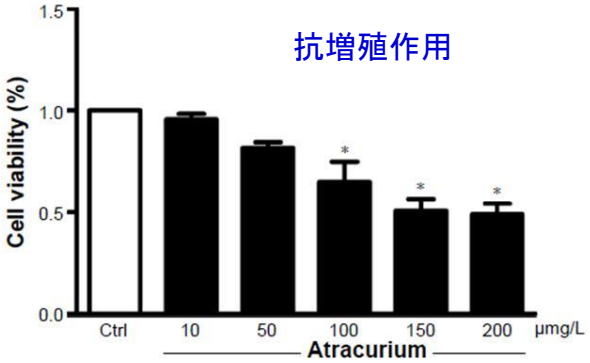
TMDU

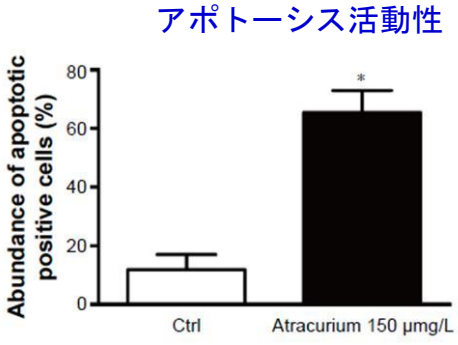
タンパク質相互作用ネットワーク DR 結果の検証

Drug ID	Drug name	Target	Score	Rank
DB00416	Metocurine Iodide	CHRNA2	0.966581	1
DB00565	Cisatracurium besylate	CHRNA2	0.966581	1
DB00732	Atracurium	CHRNA2	0.966581	1
DB00657	Mecamylamine	CHRNA2	0.966581	1
DB02457	Undecyl-phosphinic acid butyl ester	LIPF	0.953846	5

- **HPRDとSTRINGS**の両方のランダム歩行で**145薬剤・化合物**が共通
- 最高スコアを挙げた**Atractium**を選択
- 標的は**CHRNA2**(Cholinergic Receptor Nicotinic Alpha 2）**でアポトーシス経路である**
- **培養細胞A549**（ヒト肺胞基底上皮腺癌細胞）の抗増殖作用を確認



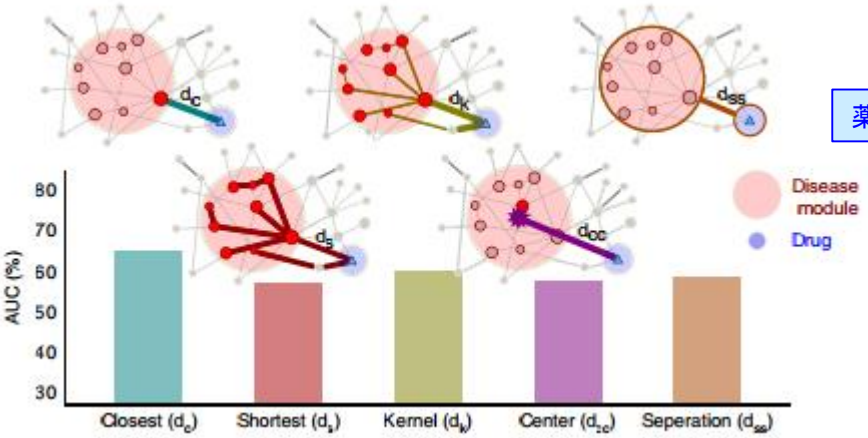




TMDU

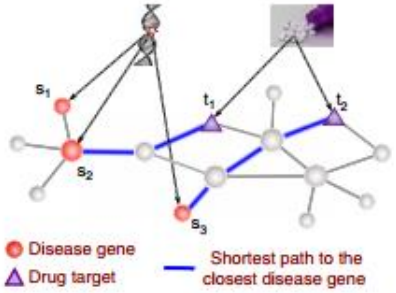
タンパク質相互作用ネットワークでの近接性によるDR

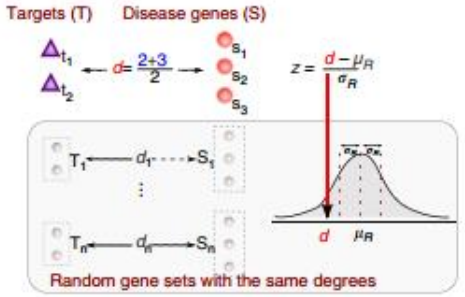
相対近接指標d_c：
①最近接の疾病関連分子との最短経路長の平均
②同じサイズで度数の分布より近接指標を計算して規格化⇒zスコア
（z < - 0.15⇒近接）
②様々な近接指標の中ではclosest measure d_c が一番薬効を予測する

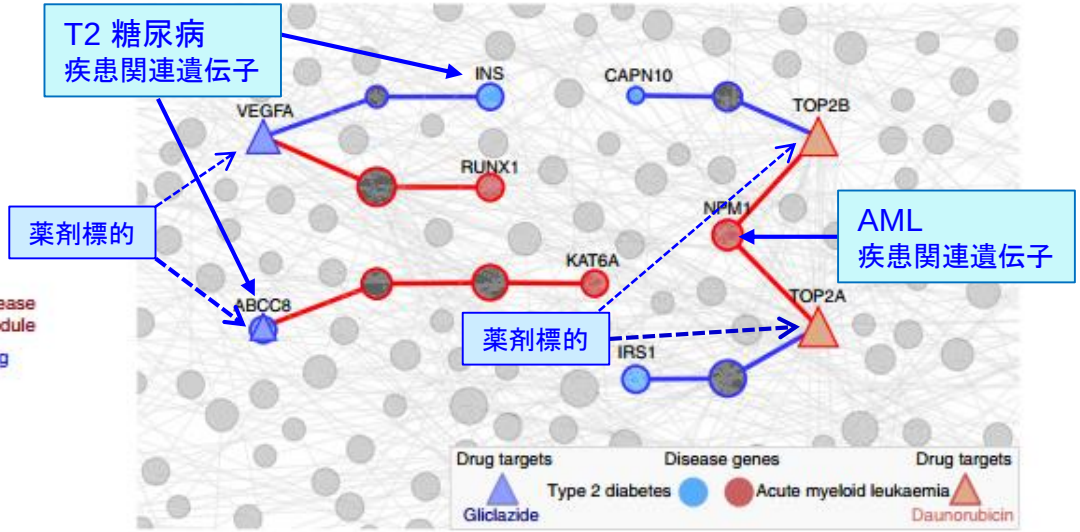


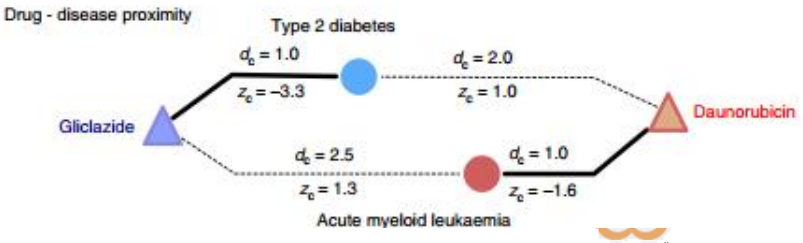
大半の薬剤は標的と疾患関連分子
２リンク離れている

(Guney, Barabasi,2016, Nat. Com)



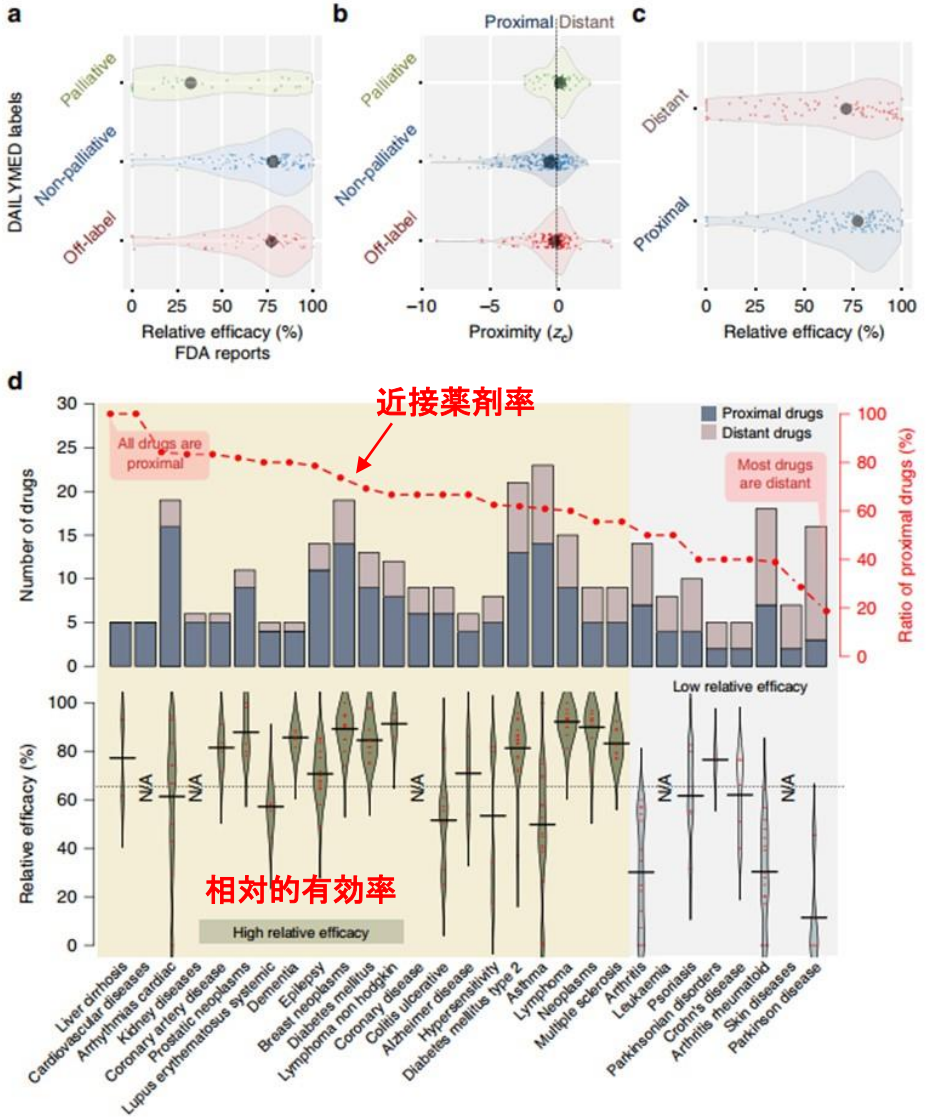






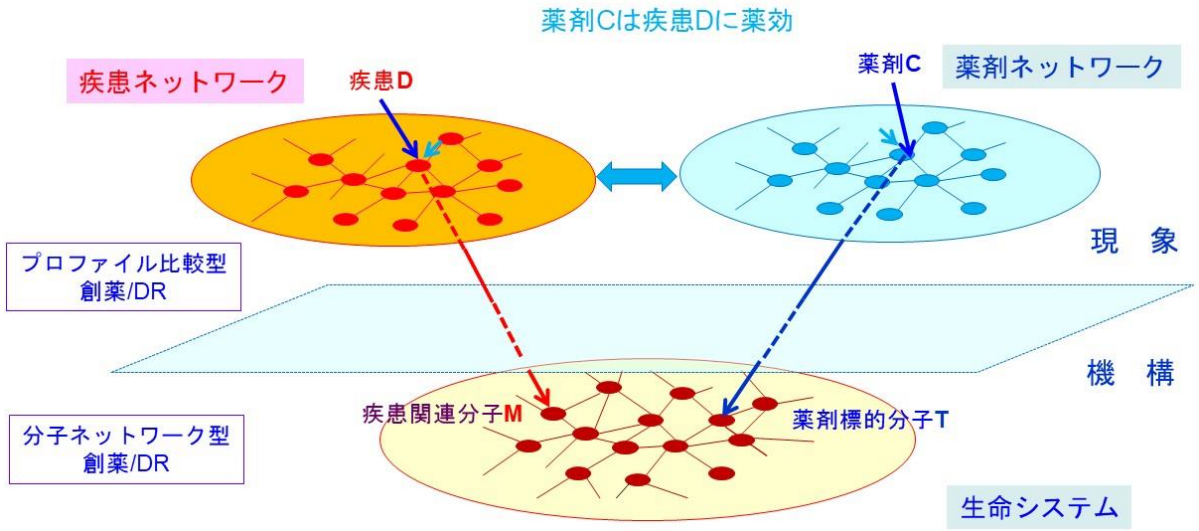
相対近接性による薬効予測

- 疾患モジュールの内部/近接に標的分子を持つ必要がある
- これまでの研究では疾患関連分子と標的分子の距離が大
  - 対症療法・緩和療法：疾患原因ではなく症状を標的としている
  - 標的分子が疾患関連分子の数は少ない（402対のうち62）
- 既成の薬は疾患と近接的である
- 緩和療法は遠隔的である
- Off-labelは緩和より近接的である
- 近接薬剤の治験の頻度は高い
- 薬剤は選択的であるが排他的ではない
- 相対的有効性と近接指標は相関する
- 平均の標的分子の数は3.5個である



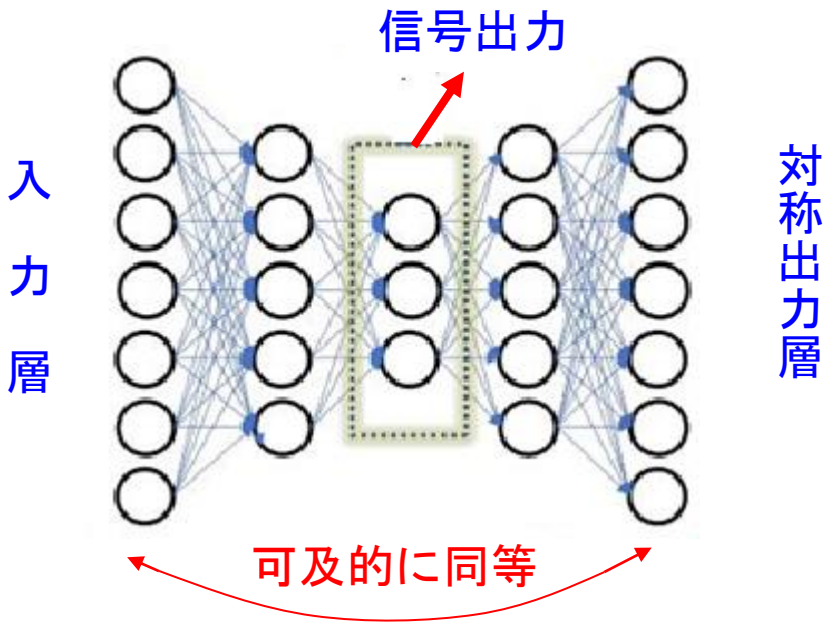
タンパク質相互作用ネットワークでの疾患-薬剤-標的分子の関係性の学習

- ビッグデータ創薬/DR
  - タンパク質相互作用ネットワーク上での有効性予測
  - 基準指標：「疾患関連分子」と「薬剤標的分子」の距離
  - 判定情報量が不足
- AI創薬/DR
  - ビッグデータ創薬/DRの限界をAI学習で補完
    - 既成の疾患-薬剤-標的分子の正例を学習（DrugBank)
    - 疾患関連分子と標的分子のPPIN位置関係性をDLで学習
  - 学習された関係性より各分子の標的分子の有効性を判定



DLの革命点 Autoencoder

- 自己符号化器を多層に構成する
  - 積層自己符号化器（stacked autoencoder）
- 入力層と出力層を対称に層構成する
  - 深層自己符号化器（deep autoencoder）

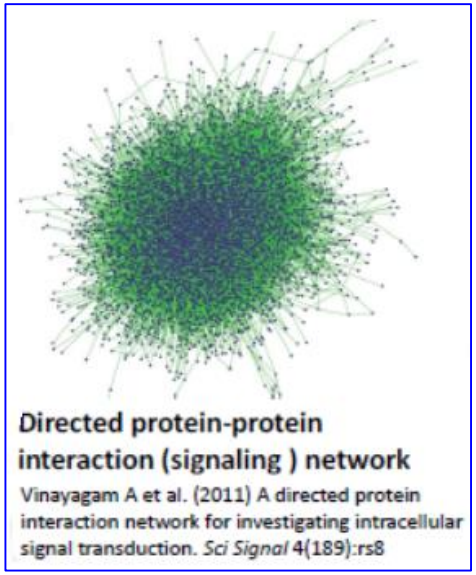


特徴的ネットワーク基底への分解

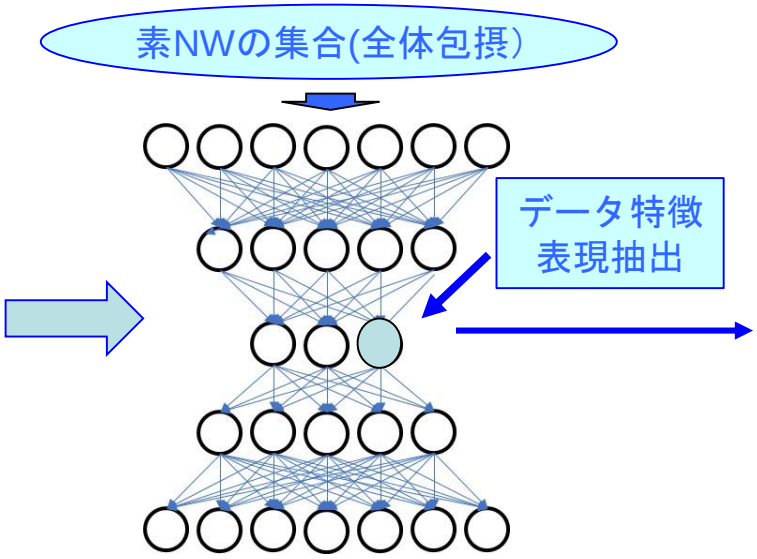
特徴的ネットワーク基底の和に縮約

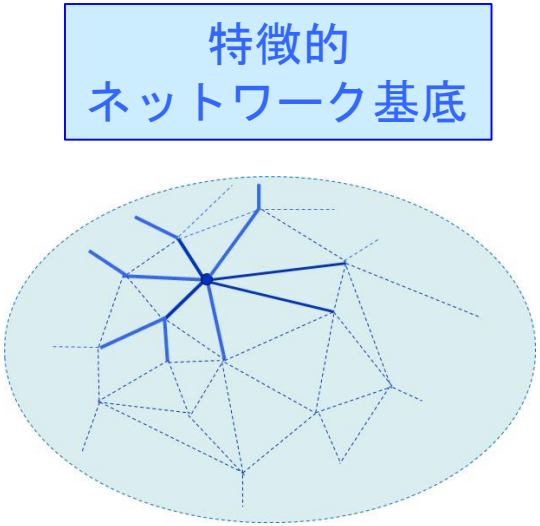
特定のノードを起点とした素NW（部分NW）の集合
全体NWを包摂する集合にDL反復自己学習

特徴的ネットワーク基底：トポロジーのみの構造/頻度構造



PPIネットワーク





Deep Learningによる創薬・DR

１）生体ネットワーク（PPIN) 特徴量の抽出
- **タンパク質相互作用ネットワーク(PPIN)**のNW結合を学習し特徴表現（特徴NW基底）を出力。
- 学習集合を部分ネットワークの集合から決める
- ノードを起点とした素NWでPPIN全体を覆う集合

２）多層Deep Auto-encoderのDLで学習.
- 特徴的NW基底の「教師無し」学習
- 次元縮約による特徴的NW基底の抽出

３）DL特徴NW基底空間における正例補完
- DrugBankからの正例とその増加（SMOTE法）

４）DL特徴NW基底量を用いた機械学習分類
- Xgboot法などを用いたDL特徴量からの判別

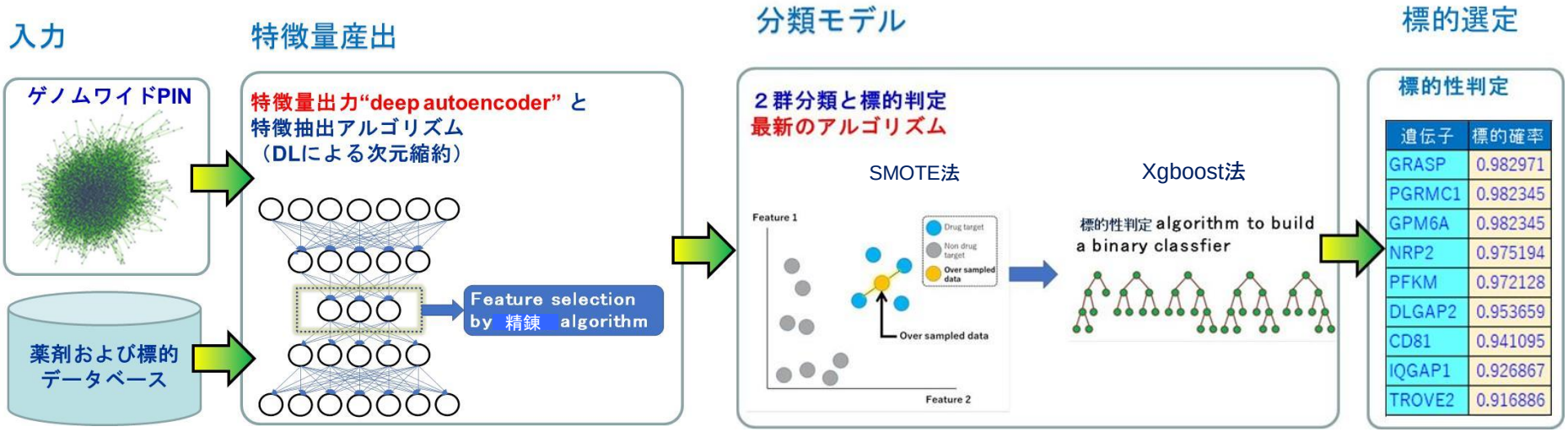
ネットワーク・タンパク質の標的性の判定

TMDU

Deep Learningによる創薬・DR

分類部　**DrugBankを利用した**
当該分子を標的とする既製薬剤の探索

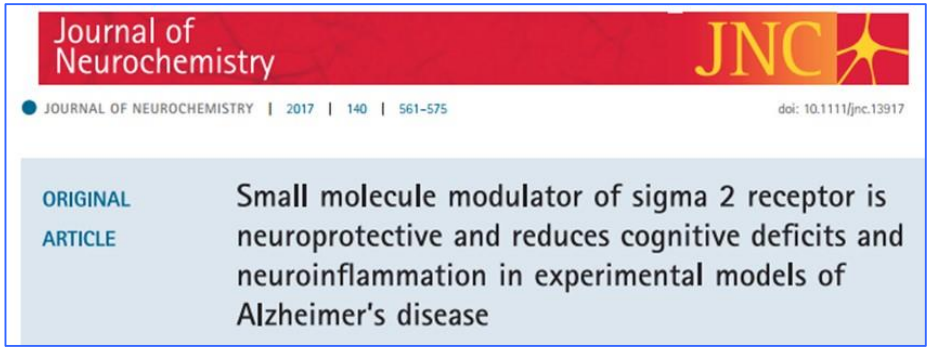
既製薬剤がない→新規薬剤探求（創薬）
既製薬剤がある→DRの検討



従来の機械学習（Random Forrest)と同じ成果は得られている

実験的研究との付合 1

PGCM1：progesterone receptor membrane 1

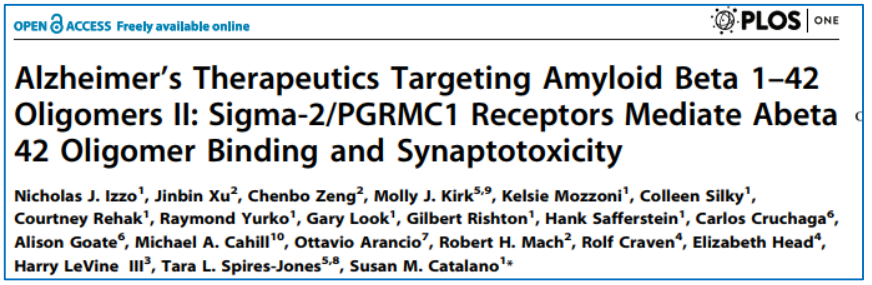

Journal of Neurochemistry

JNC

JOURNAL OF NEUROCHEMISTRY | 2017 | 140 | 561–575 doi: 10.1111/jnc.13917

ORIGINAL ARTICLE

Small molecule modulator of sigma 2 receptor is neuroprotective and reduces cognitive deficits and neuroinflammation in experimental models of Alzheimer's disease

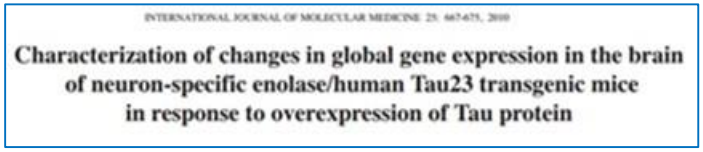

OPEN ACCESS Freely available online

PLOS ONE

Alzheimer's Therapeutics Targeting Amyloid Beta 1–42 Oligomers II: Sigma-2/PGRMC1 Receptors Mediate Abeta 42 Oligomer Binding and Synaptotoxicity

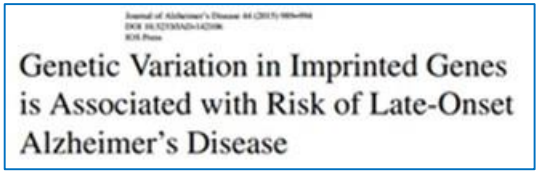
Nicholas J. Izzo[1], Jinbin Xu[2], Chenbo Zeng[2], Molly J. Kirk[5,9], Kelsie Mozzoni[1], Colleen Silky[1], Courtney Rehak[1], Raymond Yurko[1], Gary Look[1], Gilbert Rishton[1], Hank Safferstein[1], Carlos Cruchaga[6], Alison Goate[6], Michael A. Cahill[10], Ottavio Arancio[7], Robert H. Mach[2], Rolf Craven[4], Elizabeth Head[4], Harry LeVine III[3], Tara L. Spires-Jones[5,8], Susan M. Catalano[1]*

GPM6A：Glycoprotein M6A

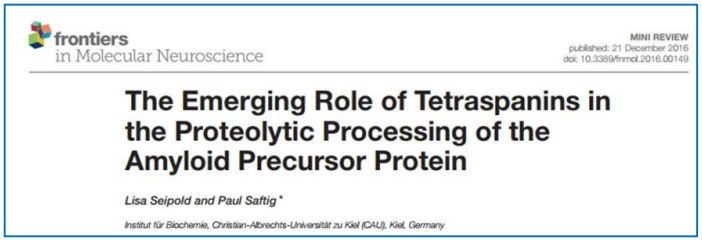

INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE 25: 667-675, 2010

Characterization of changes in global gene expression in the brain of neuron-specific enolase/human Tau23 transgenic mice in response to overexpression of Tau protein

DLGAP2：DLG-Associated Protein 2


Genetic Variation in Imprinted Genes is Associated with Risk of Late-Onset Alzheimer's Disease

CD81:Tetraspanins family


frontiers in Molecular Neuroscience

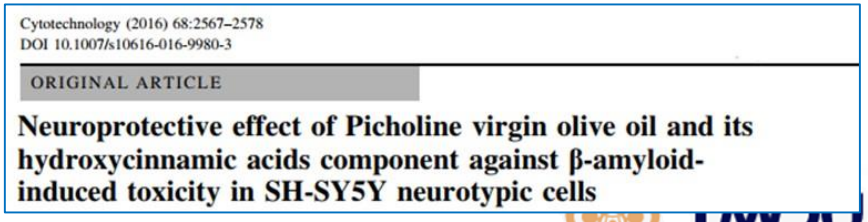
MINI REVIEW published: 21 December 2016 doi: 10.3389/fnmol.2016.00149

The Emerging Role of Tetraspanins in the Proteolytic Processing of the Amyloid Precursor Protein

Lisa Seipold and Paul Saftig*

Institut für Biochemie, Christian-Albrechts-Universität zu Kiel (CAU), Kiel, Germany

PFKM: Phospofructokinase


Cytotechnology (2016) 68:2567–2578
DOI 10.1007/s10616-016-9980-3

ORIGINAL ARTICLE

Neuroprotective effect of Picholine virgin olive oil and its hydroxycinnamic acids component against β-amyloid-induced toxicity in SH-SY5Y neurotypic cells

実験的研究との付合２

GRASP	PIK3C2B	PKIA
PGRMC1	NEU3	PFKP
GPM6A	SLC25A38	PAN2
NRP2	TNFSF12	GLUD1
PFKM	ADRA1B	DNM3
DLGAP2	DPM2	ITGA5
CD81	NLRP12	RILPL2
IQGAP1	NLRC4	MAEA
TROVE2	UIMC1	NCDN
TOP3B	IL8	DGCR14
TJP1	VAV1	PACSIN3
PDGFB	ARHGEF1	CD46
SETD2	WISP2	NIT1
CFLAR	PRKCE	ICAM4
PROS1	TBXA2R	GNA13
SIT1	TSPAN4	STK40
SIGLEC7	EPHB4	ROGDI
SHC2	LOC63920	CDH10
SH2D1A	PSEN1	WSB2
	SPOCK3	PHPT1
	TSPO	
	SLC4A1	

アルツハイマー症に対する有効な薬剤標的分子の候補を100以上見出した。

SLC25A38 (APPOPTOSIN)

SLC25A38は アルツハイマー症・脳梗塞患者の脳において増加。 さらに、SLC25A38 の発現低下はBax/BH3IやAβ/glutamateによって誘導されるニューロンの死亡によるアポトーシスを抑制する

Previous Next

Featured Article | Articles, Cellular/Molecular

Appoptosin is a Novel Pro-Apoptotic Protein and Mediates Cell Death in Neurodegeneration

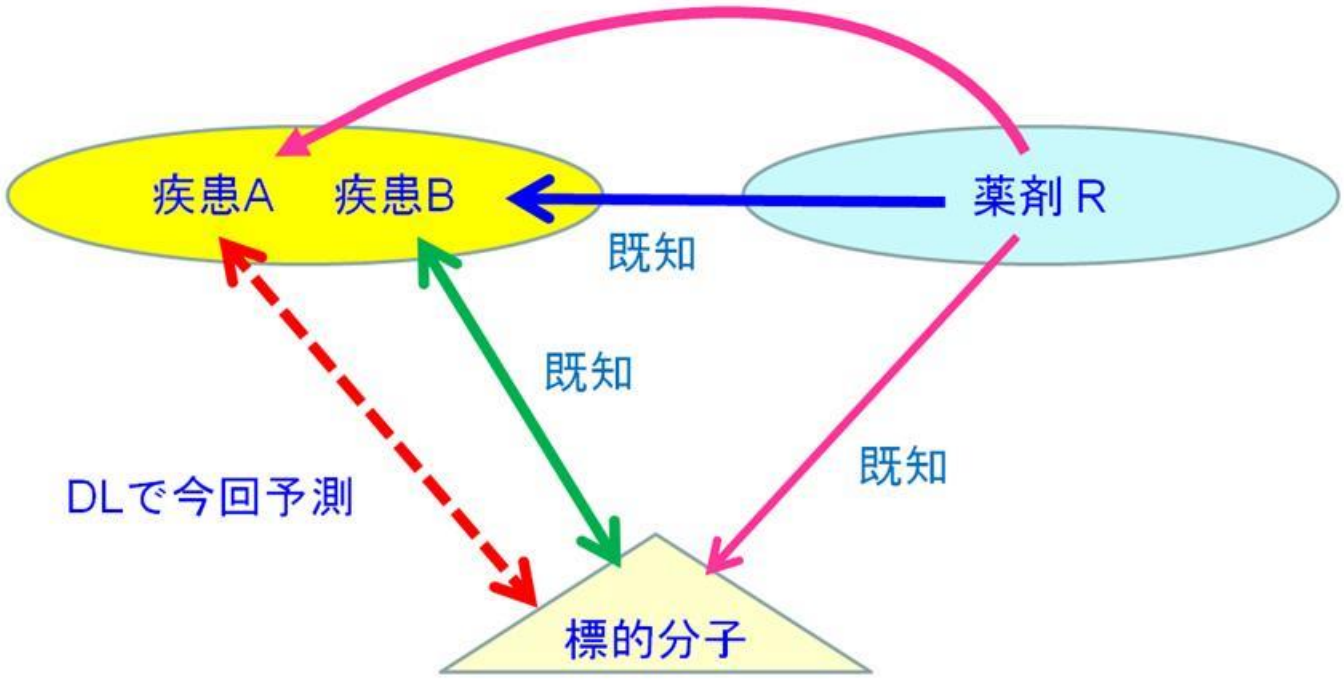
Han Zhang, Yun-wu Zhang, Yaomin Chen, Xiumei Huang, Fangfang Zhou, Weiwei Wang, Bo Xian, Xian Zhang, Eliezer Masliah, Quan Chen, Jing-Dong J. Han, Guojun Bu, John C. Reed, Francesca-Fang Liao, Ye-Guang Chen, and Huaxi Xu

Journal of Neuroscience 31 October 2012, 32 (44) 15565-15576; DOI: https://doi.org/10.1523/JNEUROSCI.3668-12.2012

TMDU

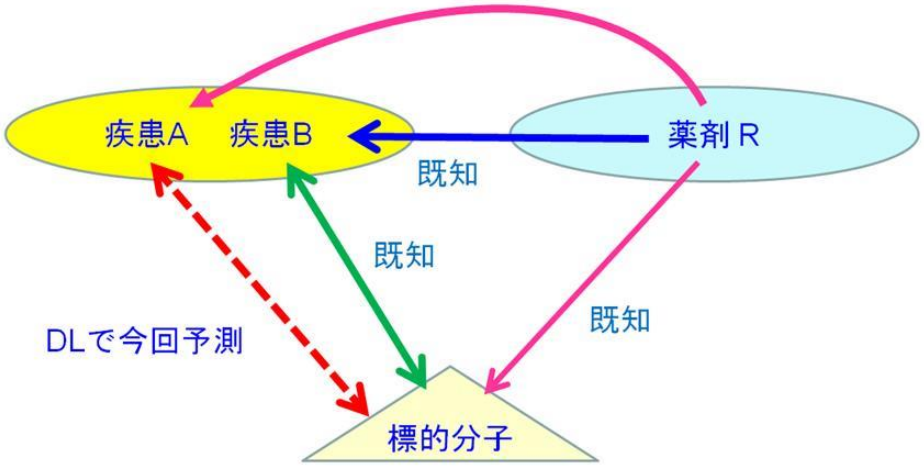
アルツハイマー症のDR薬剤候補

DR候補薬剤：アルツハイマー症の候補標的分子が、ある既承認薬剤Rの標的分子と同一あれば薬剤Rは、アルツハイマー症にも有効と期待。



アルツハイマー症のDR薬剤候補

DR候補薬剤：
アルツハイマー症の候補標的分子が、ある既承認薬剤Rの標的分子と同一であれば薬剤Rは、アルツハイマー症にも有効と期待。



repositonable drug	taregt	# of target	category
Tamoxifen	PRKCB PRKCE PRKCG ESRRG	4	Anti-Estrogens; Antineoplastic Agents; Antineoplasti
Mianserin	SLC6A4 DRD3 OPRK1 ADRA1B	4	Adrenergic Agents; Adrenergic alpha-Antagonists; A
Amitriptyline	SLC6A4 OPRK1 ADRA1B OPRM1	4	
Dextromethorphan	SLC6A4 PGRMC1 OPRM1 OPRK1	4	Alkaloids; Antitussive Agents; Central Nervous Syste
Mirtazapine	OPRK1 ADRA1B DRD3 SLC6A4	4	Adrenergic Agents; Adrenergic alpha-Antagonists; A
Tramadol	OPRM1 OPRK1 SLC6A4	3	Alcohols; Amines; Analgesics; Analgesics, Opioid; C
Zinc	MPG SERPINA1 SERPIND1	3	Acetates; Acetic Acid; Acids; Acids, Acyclic; Acids, N
Amoxapine	SLC6A4 DRD3 ADRA1B	3	Adrenergic Agents; Adrenergic Uptake Inhibitors; Al
Etorphine	OPRM1 OPRK1 OPRL1	3	Alkaloids; Analgesics; Analgesics, Opioid; Central N
Tapentadol	OPRM1 OPRK1 SLC6A4	3	Analgesics; Analgesics, Opioid; Benzene Derivatives
Loxapine	ADRA1B DRD3 SLC6A4	3	Antipsychotic Agents; Antipsychotic Agents (First Ge
Pethidine	OPRK1 OPRM1 SLC6A4	3	Acids, Heterocyclic; Adjuvants; Adjuvants, Anesthesi
Talampanel	GRIA1	1	Benzazepines; Heterocyclic Compounds; Heterocycli
Etanercept	FCGR3B	1	Amino Acids, Peptides, and Proteins; Analgesics; A
Vitamin E	PRKCB	1	Antioxidants; Benzopyrans; Chemical Actions and Us
N-[(2R)-2-benzyl-4-(hydroxyamino)-4-	LTA4H	1	
Adalimumab	FCGR3B	1	Amino Acids, Peptides, and Proteins; Anti-Inflamm
ALPHA-HYDROXYFARNESYLPHOSPH	FNTB	1	Alcohols; Fatty Alcohols; Hydrocarbons; Lipids; Orga

AI準拠DRの例

FDA承認薬剤 **adalimumab** と **etanercept** はDR候補薬剤として期待できる。これらの薬剤はTNF-α （免疫応答を肝要なサイトカイン) の抑制分子で、TNF-α の過剰発現は、特に中枢神経系に炎症を起こす。

MedGenMed
Medscape General Medicine

MedGenMed. 2006; 8(2): 25.
Published online 2006 Apr 26.
PMCID: PMC1785182

TNF-alpha Modulation for Treatment of Alzheimer's Disease: A 6-Month Pilot Study

Edward Tobinick, MD, Assistant Clinical Professor of Medicine, Hyman Gross, MD, Clinical Professor of Neurology, Alan Weinberger, MD, Associate Clinical Professor of Medicine/Rheumatology, and Hart Cohen, MD, FRCPC, Associate Clinical Professor of Medicine/Neurology

CNS Drugs
November 2016, Volume 30, Issue 11, pp 1111–1120

Treatment for Rheumatoid Arthritis and Risk of Alzheimer's Disease: A Nested Case-Control Analysis

Authors　Authors and affiliations

Richard C. Chou, Michael Kane, Sanjay Ghimire, Shiva Gautam, Jiang Gui

TMDU

第２世代の
ゲノム・オミックス医療

ゲノム医療の第2世代

成功した臨床実装

1. **希少先天遺伝疾患**の**原因遺伝子**を**病院の現場で**シーケンサにより同定
2. **がんのドライバー遺伝子変異**を同定、適切な**分子標的薬**を処方
3. 患者の**薬剤の代謝酵素の多型性**を**先制的に同定**し、副作用を防ぐ

しかし

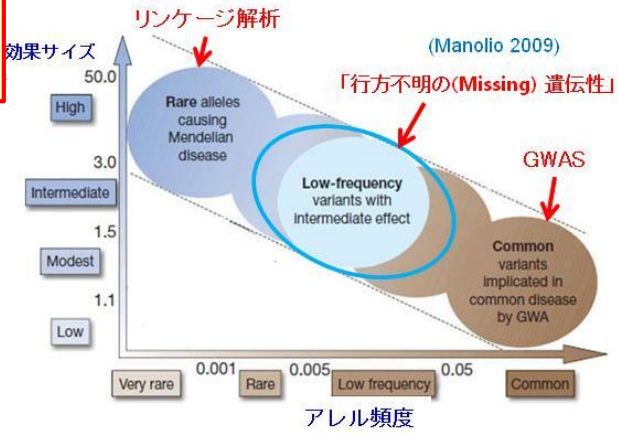
多因子疾患の機序/発症予測は無着手である

- 「単一遺伝的原因」帰着アプローチの限界
- 「行方不明の遺伝力」の主要な原因
  - 複数の疾患関連遺伝子間の相互作用: G x G
  - 環境と遺伝子の相互作用が：G x E

SNPの相対リスク
低い(1.1～1.3)理由
GxE組合せ特異的効果を環境要因の平均

多因子疾患は個人の＜遺伝的体質と環境要因＞の＜相互作用の結果。シーケンスだけでは解明不能

疾患発症の遺伝要因と環境要因の相互作用は
加算的 $(G \oplus E)$でもなく乗算的 $(G \otimes E)$ でもない
＜(G,E) 組合せ特異的な効果＞である
例　大腸がんの遺伝要因と環境（生活習慣）要因



大半の疾患の基礎としての「遺伝素因X環境要因」の相互作用

一部の単一遺伝病を除き、大半の疾患（Common diseases)の発症は

疾患発症の相対リスク=

遺伝要因(G:genome) X 環境要因(E:exposome)

相互作用は加算的 でもなく乗算的 でもない

＜**(G,E) 組合せ特異的な効果**＞である

GWASでSNPの相対リスクが低い(1.1～1.3)理由： **GxE組合せ特異的効果を環境要因の全てに亘って平均しているからである**

発達プログラム説 DOHaD

Developmental Origin of Health and Disease)

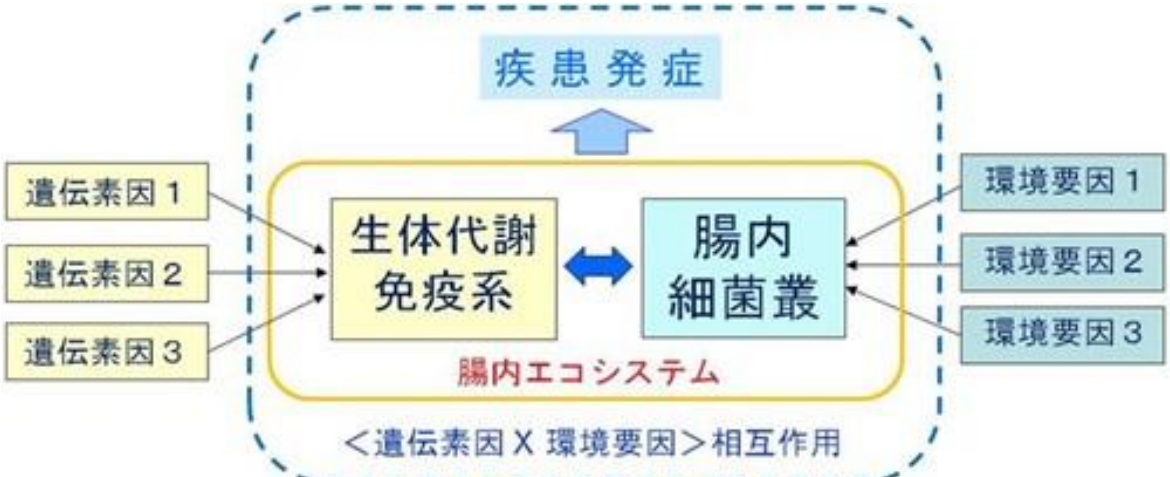
- オランダ飢饉
  - 第2次大戦末期、ナチスの封鎖、約半年間酷い飢饉
  - 飢饉の期間に胎児、戦後30年
  - 成人期:肥満,糖尿病,心筋梗塞,統合失調
- Baker仮説：英国心筋梗塞増加
- エピジェネティク機構
  - 過度な低栄養：肝臓のPPARα/γ（倹約遺伝子）メチル化低下・遺伝子発現がオン
  - エピジェネティック変化は**可変**：短期的変化、**長期的「記憶」**次の世代も

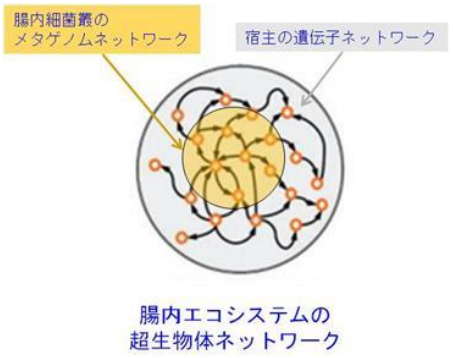
オランダ
飢饉 (1944)

環境因子 → Epigenome変化 → 遺伝子発現調節 → 疾病発症

腸内細菌叢microbiome：メタゲノム

- **疾患の環境発症要因**（exposome）
  - 腸内microbiome：環境要因の最大の１つ
- **腸管微生物叢**（gut microbiome）
  - 約1000種類、100兆個、総重量１～1.5kg,「**実質的な臓器**」
  - 遺伝子数個人あたり**約50万遺伝子**、総数：数100万遺伝子
- **免疫系、炎症系、粘膜免疫細胞群との相互作用**
  - **食物の難消化性の食物繊維**：腸内細菌によって嫌気的に代謝、酪酸などの「**短鎖脂肪酸**」がエネルギー源となる
  - 食事・栄養物質による環境要因は、腸内細菌叢の代謝物（短鎖脂肪酸やTMAOなど）から宿主の生体機構に相互作用

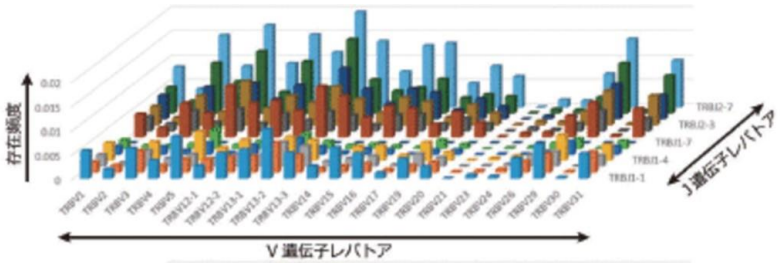




メタゲノム
超生物ネットワーク

免疫ゲノム

- 可変領域や相補性決定領域（特にCDR３）のDNAやRNAを次世代シーケンサ(HTS)で解析
- レパトア解析
  - 抗原受容体全体のプロファイルを俯瞰的に把握できる
  - V(D)Jなどの成分を基軸として3次元表示可能。
  - 疾患罹患ともに瞬時に全体像が変化する。
  - 網羅的病態全体像を提示する
    - VDJの使用頻度
    - 多様性(diversity)の変化
    - 疾病/加齢レパトア分布変化
- 臨床シーケンスに含まれる
- ３次元分布の特徴分析



（レパトア・ジェネシス社）

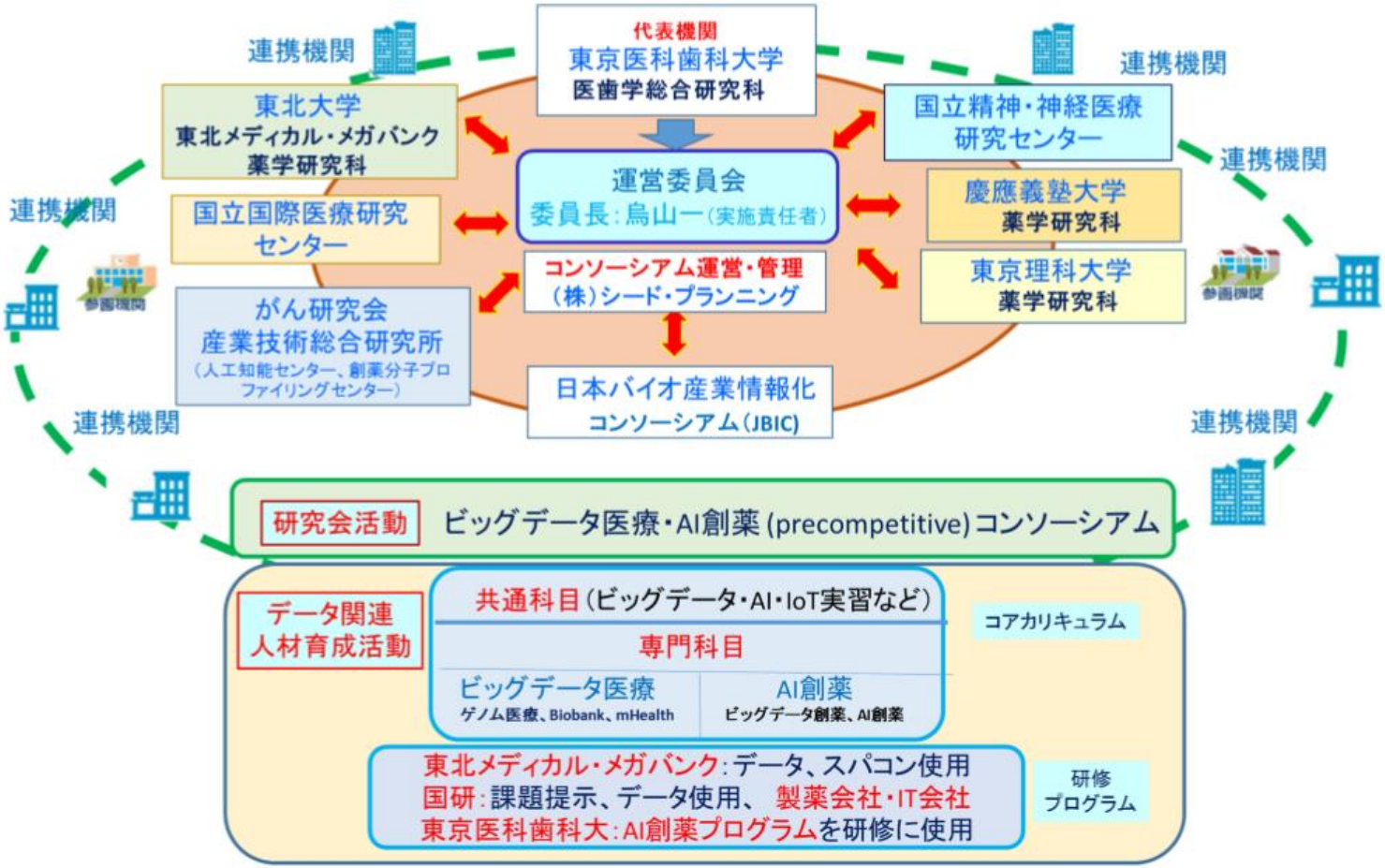
第2世代のゲノム・オミックス医療

- 生涯的全体性においてその個人の疾患可能性の全体性を把握し、個別化予防、個別化治療に取り組む
- ゲノム・オミックス情報と医療・健康
  - **Clinical Sequencingのインパクト**
- 第１世代ゲノム医療
  - ゲノムの変異・多型性の個別性に基づく
- 第２世代のゲノム医療
  - 多因子疾患が対象、環境情報との相互作用
  - エピゲノム、メタゲノム・免疫ゲノムなど

疾患メタ・オミックス修飾

TMDU

ビッグデータ医療・ＡＩ創薬コンソーシアム



2018.2.23-25, Harvard/MIT/TMDU - Datathon

TMDU

ご清聴ありがとうございました